IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 10, 2020, accepted March 3, 2020, date of publication March 6, 2020, date of current version March 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978875

Radar Signal Modulation Recognition Based on Deep Joint Learning

DONGJIN LI, RUIJUAN YANG, XIAOBAI LI, AND SHENGKUN ZHU
Air Force Early Warning Academy, Wuhan 430019, China

Corresponding author: Dongjin Li (li_dong_jin@163.com)

This work was supported in part by the National Defense Science and Technology Innovation Zone Fund of China under Grant 17H86304ZT00302201.

ABSTRACT The development of integrated avionics systems and electromagnetic spectrum technology has attracted widespread attention. It has further increased the performance requirements for modulation signal recognition technology in complex electromagnetic environments. Therefore, this paper proposes a deep joint learning technique, including deep representation and low-dimensionality discrimination, to enhance feature stability and environmental adaptability. Specifically, we design a feature learning network based on AlexNet to extract in-depth features and optimize it through parameter-based transfer learning techniques, promote multi-level representation capabilities of features and reduce the sample size requirements. Moreover, we propose a classification algorithm based on kernel collaborative representation and discriminative projection to enhance the ability of low-dimensionality representation and between-class discrimination, which optimized using the mini-batch random gradient descent method. As shown in the simulation, the overall average recognition success rate of this method aiming at twelve radar signal modulation types reaches 97.58% at SNR of −6dB. The results of simulation and analysis demonstrate the superiority of the proposed model in terms of robustness, timeliness, and adaptability to small samples.

INDEX TERMS Radar signal modulation recognition, deep representation, kernel collaborative representation, discriminative projection.

I. INTRODUCTION

Fast and efficient radar signal modulation recognition plays a critical role in the electromagnetic frequency spectrum domain confrontation. It directly affects the accuracy of subsequent tasks, such as electronic countermeasures [1]. The traditional technology of modulation recognition relies on feature engineering and classifier models, such as the pulse description word (PDW) based recognition method [2], artificial feature extraction and classification [3]–[5], and the combined approach of nonlinear feature extraction and machine learning classification [6]–[11]. These traditional methods have been widely applied in this field and have obtained many achievements. However, it is challenging to adapt to multiple types of signals and complex electromagnetic environments because of its deficiency on feature deep representation. In recent years, many scholars have studied modulation recognition technology based on the representation learning paradigm, including dictionary learning and deep learning [12]–[15]. In [16]–[19], different deep network structures were designed for classification and other tasks, including a structured autoencoder and multi-view linear discriminant analysis networks. As is known, deep learning has stronger deep representation capabilities and classification performance because of its deep mapping mechanism. Recognition algorithms based on deep networks and time-frequency features have achieved excellent performance. In [20], the recognition method based on a deep belief network (DBN) and short-time Fourier transform (STFT) achieved a recognition success rate (RSR) of 95% when the SNR is −5dB. In [21], the LeNet-5 based convolution neural network (CNN) reaches an RSR of 93.7% for eight types of radar signals when the SNR is −2dB. In [22]–[28], pre-processing algorithms such as the sample averaging technique, image noise reduction technique, and convolutional denoising autoencoder enhanced the stability of time-frequency features; and different CNNs improved the recognition performance. However, these methods still

The associate editor coordinating the review of this manuscript and approving it for publication was Xi Peng.

depend on prior knowledge and feature enhancement processing to some extent. In addition, deep learning requires high computing costs and a large number of samples, so its advantages can hardly be realized in cases of insufficient samples. In [29], [30], deep learning-based transfer learning algorithms and feature extraction algorithms were designed to obtain deep features for subsequent recognition, thereby reducing the sample size requirements of deep networks. However, the high-dimensionality deep features learned from transfer learning networks have poor interpretability, and full connection (FC) layers of deep networks usually introduce considerable parameters and large computational overheads, which is not conducive to effective recognition. In addition, the adaptability of deep migration networks to new samples can be further improved through auxiliary classifiers [31]. Therefore, a well-designed joint recognition structure will improve the low-dimensionality representation of features, the computational complexity of FC layers, and adaptability with small samples.

Dictionary learning (DL) has been widely applied in fields such as cognitive radio, and image identification, which can realize data-based essential feature learning and discrimination with low-dimensionality representation. Dimensionality reduction (DR) or low-dimensionality representation is also an important issue for dictionary classification. The performance of dictionary learning is strongly dependent on the choice and optimization of various discriminant items, such as regular constraints and discriminant constraints. In [32], regular constraints of l_1-norm and $l_{1,2}$-norm strengthened the discriminative learning ability, whereas the sparsity and computational complexity limited their applications. Therefore, a collaborative representation classification method with an l_2-norm regular constraint was proposed in [33], which is very efficient and time-saving. In [34]–[44], discriminant constraint models such as the label consistent K-SVD (LC-KSVD) dictionary, low-rank shared dictionary, multilayer convolutional sparse dictionary, scalable block-diagonal locality-constrained projective dictionary, and structured analysis discriminative dictionary, were designed to simultaneously enhance classification performance and reduce the computational cost. However, these methods perform DR and DL independently, which may result in not fully exploiting the discriminative information of the training data. In [45]–[47], a joint discriminative dimensionality reduction and dictionary learning (JDDRDL) model, a simultaneous dimensionality reduction and dictionary learning (SDRDL) model, and a dictionary learning for sparse representation-based classification (DSRC) model were proposed for simultaneous learning of low-dimensionality representation and classification dictionary. These models are suitable for linearly separable features. In [48]–[50], the kernel methods were used to enhance the nonlinear representation ability, which improved the linear inseparability of features. In [51], a kernel dictionary learning-based discriminant analysis (KDL-DA) method was proposed to learn low-dimensionality representation and a classification dictionary in kernel space. However, the KDL-DA easily falls into a locally optimal solution. Therefore, we further combine the method of discriminative projection and kernel collaborative representation to enhance the between-class differences and timeliness of dictionary classification. Although dictionary learning methods are diverse, their deep representation capabilities are limited. Moreover, deep learning and dictionary learning have advantages in deep representation and low-dimensionality representation of data, respectively. Therefore, this paper combines these two representative learning paradigms for learning the low-dimensionality discriminative representation in deep architecture and improving classification performance.

In this paper, the deep joint learning-based modulation recognition algorithm is exploited. It aims to improve RSR under small samples and a low signal-to-noise ratio (SNR) environment. The overall performance of the proposed method is verified for twelve radar signal modulation types, including BPSK, LFM, Costas, Frank, P1-P4 codes, and T1-T4 codes, under a wide range of SNR. Our major contributions are summarized as follows. (1) Design a CNN-based feature learning network, which is initialized by AlexNet to extract in-depth features with higher stability for recognition. (2) Propose a classification technique based on kernel collaborative representation and discriminative projection (KCRDP), in which the objective function is constructed by the Fisher criterion and optimized by a mini-batch stochastic gradient descent (MSGD) method with dynamic increment constraints. The timeliness and stable low-dimensionality representation ability of the model are emphasized. (3) Propose a deep joint learning algorithm based on CNN and KCRDP (CNN-KCRDP), including deep learning and kernel dictionary learning, which improves the adaptability of small samples and reduces the computational complexity without prior knowledge and feature enhancement processing.

The paper is organized as follows. The modulation recognition system of the radar signal is introduced in Section II. Section III describes the signal model and the pre-processing method. Section IV describes the details of the deep joint learning framework, including the deep network for feature learning, classification model, and optimization method. Section V conducts a simulation comparison and makes discussions. Finally, the conclusions are drawn in Section VI.

II. SYSTEM OVERVIEW

The proposed modulation recognition system consists of two parts: pre-processing and recognition. The details are shown in Fig.1. First, intercepted intra-pulse signals are transformed into time-frequency images (TFIs) by Choi-Williams distribution (CWD) without feature enhancement processing, and the TFI size is adjusted to fit CNN-KCRDP. Second, the TFIs are divided into a training set and a testing set, which are used for CNN-KCRDP optimization and classification testing, respectively. Moreover, the structure of CNN-KCRDP is shown in Fig.2, which includes the convolutional layers and equivalent FC layers (named KCRDP layers).

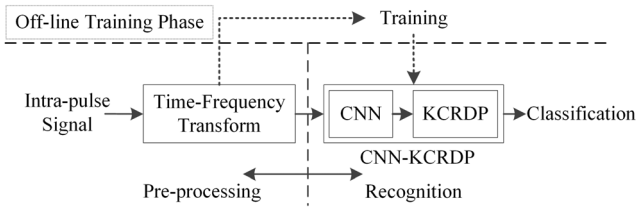

FIGURE 1. The proposed modulation recognition system of the radar signal.

The convolutional layers are initialized through transfer learning and fine-tuned through the training set to extract in-depth nonlinear features. The convolution, pooling, and normalization processes in the convolutional layers make it robust to the geometrical distortions. In addition, KCRDP layers are optimized through MSGD based kernel discriminative dictionary learning to reduce the feature dimensionality. Specifically, the KCRDP layers are divided into three parts: a kernel mapping for enhancing feature separability, a collaborative representation for improving the timeliness of classification, and a discriminative projection for forming low-dimensionality representation.

III. SIGNAL MODEL AND PRE-PROCESSING

A. SIGNAL MODEL

In the electromagnetic environment, the intercepted single-pulse signal is disturbed by additive white Gaussian noise (AWGN). The sampling sequence model is given by

$$x(k) = \mathrm{A}\exp(\mathrm{j}(2\pi f_0 k + \varphi(k) + \theta_0)) + n(k) \quad (1)$$

where k is an integer, and A is amplitude. f_0 is the initial frequency, θ_0 is the initial phase, $\varphi(\cdot)$ is the instantaneous phase, which implies the intra-pulse modulation information, and $n(\cdot)$ is the Gaussian white noise. We mainly consider twelve types of radar signals, including frequency modulated signals (such as LFM, Costas), phase-modulated signals (BPSK), polyphase codes (such as Frank, P1, P2, P3, and P4), and polytime codes (such as T1, T2, T3, and T4) [52]. The instantaneous phase of the radar signal is shown in Table 1.

TABLE 1. The instantaneous phase representation of twelve radar signal types considered in this paper.

Modulation type	$\varphi(k)$
BPSK	0 or π
LFM	$k^2B/2\tau_{pw}$
Costas	$2\pi f_j k$
Frank	$2\pi(i-1)(j-1)/M$ $i=j=1,2,3,\ldots,M$
P1	$-\pi[M-(2j-1)][(j-1)M+(i-1)]/M$ $i=j=1,2,3,\ldots,M$
P2	$-\pi[2i-1-M][2j-1-M]/2M$ $i=j=1,2,3,\ldots,M$
P3	$\pi(i-1)^2/N_c, i=1,2,\ldots,N_c$
P4	$\pi(i-1)^2/N_c-\pi(i-1), i=1,2,\ldots,N_c$
T1	$\mathrm{mod}\{2\pi\mathrm{INT}[(mk-jT)jn/T]/n,2\pi\}$ $j=0,1,2,\ldots,k-1$
T2	$\mathrm{mod}\{2\pi\mathrm{INT}[(mk-jT)(2j-m+1)n/2T]/n,2\pi\}$ $j=0,1,2,\ldots,k-1$
T3	$\mathrm{mod}\{2\pi\mathrm{INT}[(n\triangle Fk^2)/2t_m]/n,2\pi\}$
T4	$\mathrm{mod}\{2\pi\mathrm{INT}[(n\triangle Fk^2)/2t_m-(n\triangle Ft)/2]/n,2\pi\}$

B. CHOI-WILLIAMS DISTRIBUTION AND TIME-FREQUENCY ANALYSIS

The time-frequency transform of the Choi-Williams distribution (CWD) has a relatively stable characterization ability for local characteristics and overall structure. The CWD can be expressed as

$$C(\omega,t) = \iiint_{\infty} f(\eta,\tau)e^{j2\pi\eta(s-t)}x(s+\tau/2)x^*(s-\tau/2)e^{j\omega\tau}\,d\eta\,ds\,d\tau \quad (2)$$

where $C(\omega, t)$ is the time-frequency distribution function, ω and t refer to coordinates of time and frequency, $f(\eta, \tau)$ is a kernel function and satisfies $f(\eta,\tau) = e^{(\pi\eta\tau)^2/2\delta}$, δ is the controllable factor to suppress the cross terms, and cross terms decrease with δ. δ is set to 1 balanced resolution and cross terms.

The time-frequency images of the CWD shown in Fig.3 have good time-frequency structure and high-resolution details. In addition, polyphase codes (P1 and P4, Frank and P3) and polytime codes (T1 and T3, T2 and T4) have

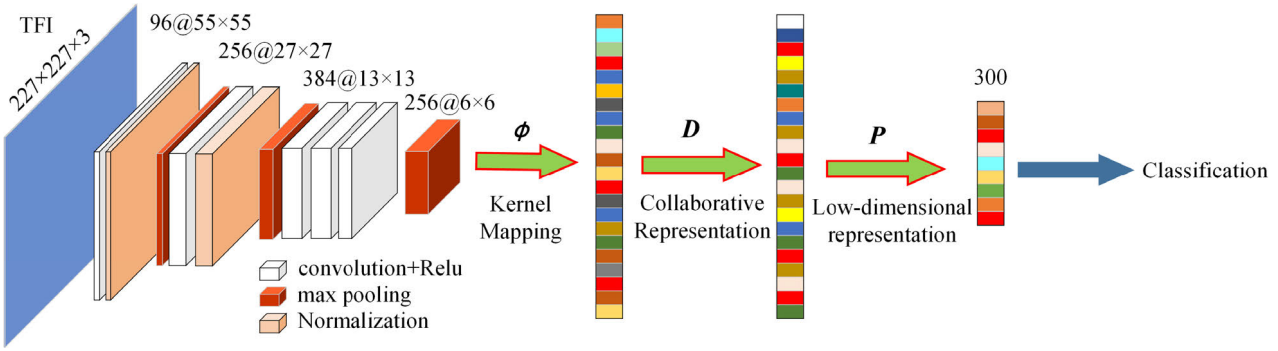

FIGURE 2. An example using the proposed deep joint learning algorithm for classification. The primary processing of CNN-KCRDP includes deep feature extraction, kernel mapping, collaborative representation, low-dimensionality representation, and classification.

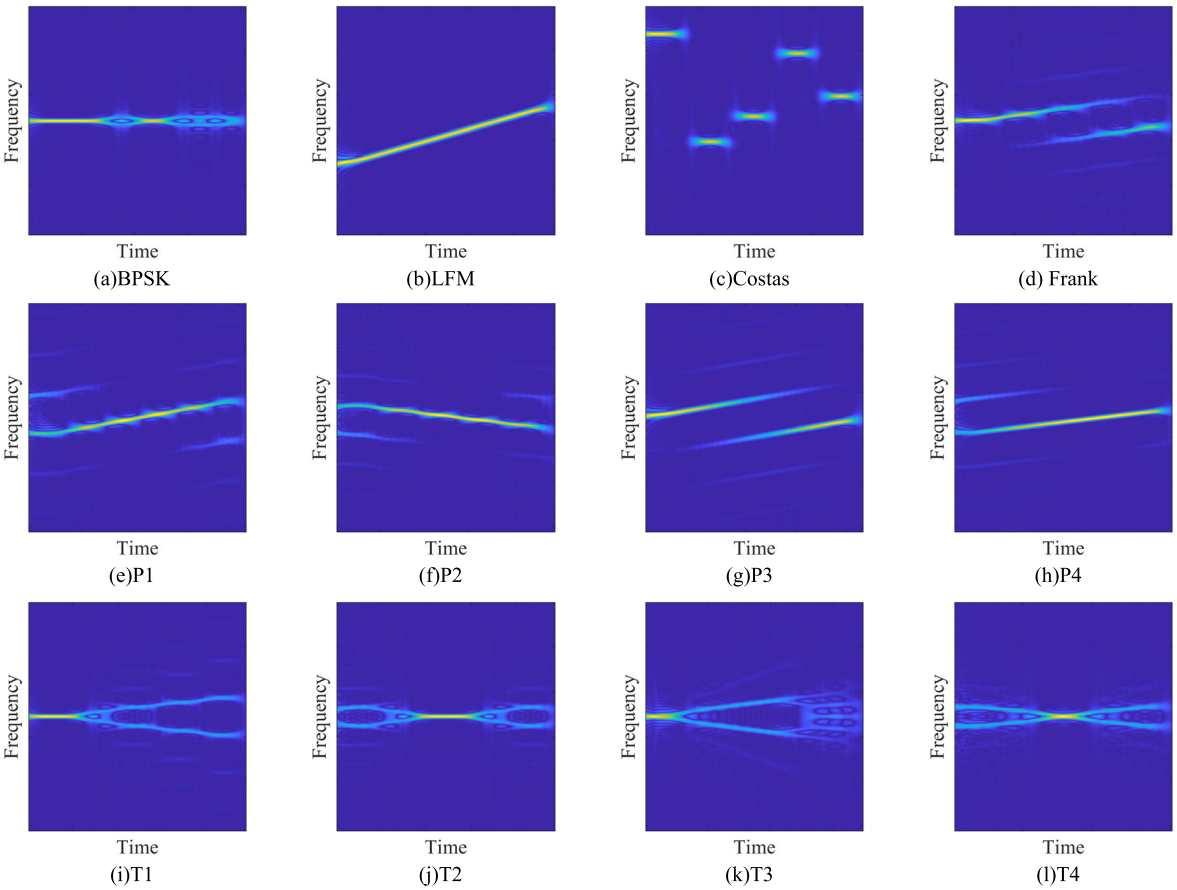

FIGURE 3. The image displays the time-frequency distribution (TFD) of twelve radar signals at SNR of 30dB.

higher similarity in structure. Feature enhancement algorithms such as noise reduction and dimensionality reduction [15], [20]–[22], [28], [30] have been widely used in radar signal recognition, which reduces the need for classifier capabilities. However, these algorithms require more expertise and may introduce additional computational complexity and a certain degree of information loss. In this paper, the CWD feature is used as the initial feature without any other feature enhancement processing, and it is adjusted to a suitable size for the network input by a bicubic interpolation algorithm [53].

IV. THE PROPOSED DEEP JOINT LEARNING FRAMEWORK

In this section, the deep joint learning framework is given, which combines the deep representation advantages of CNN and the discrimination advantages of kernel dictionary learning. Specifically, CNN is used to extract in-depth features, and KCRDP is used to achieve low-dimensionality representation and efficient recognition.

A. CNN-BASED FEATURE LEARNING

The parameters of deep neural networks are mainly included in the FC layers, whereas the parameters of the convolutional layers are relatively small and time-efficient. In addition, the convolutional layers have the characteristics of parameter sharing and translation invariance. These well-known transmission capabilities of convolutional layers have been verified by various classic CNNs, such as AlexNet, GoogleNet, and VGG-net. In our frameworks, the convolutional layers have the same structures as the corresponding baseline models, whereas all FC layers after the final pooling layer are replaced with the KCRDP layers. In general, problems such as poor convergence and reduced timeliness increase as the network deepens in the case of small samples. It is more reasonable to choose a medium-sized network and use transfer learning technology for optimization [54]. In this paper, AlexNet is selected for feature learning. Then, the CNN parameters are initialized by performing pre-training and fine-tuning using its original network. Specifically, the network can be pre-trained through the ImageNet database to reduce sample capacity requirements caused by random initialization; and fine-tuned through the TFI dataset to enhance the adaptability of the radar dataset further. The deep features extracted from the convolutional layer will be more stable than CWD features.

Compared with the conventional method, the proposed feature learning network has the following advantages. First, the network is optimized through parameter-based transfer

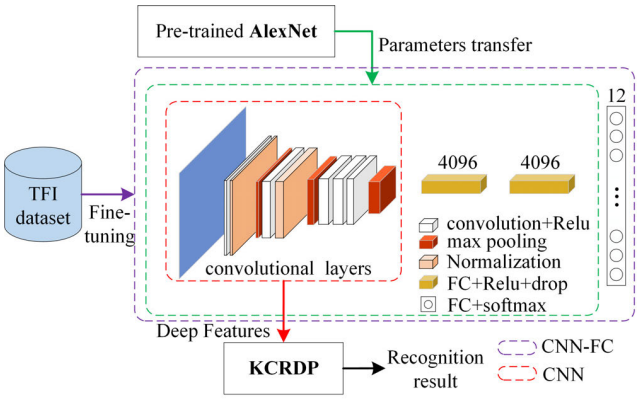

FIGURE 4. The structure and optimization process of the feature learning network based on CNN.

learning, which reduces the sample capacity requirements and the computational overhead caused by random initialization. Second, the features are extracted using convolutional layers, which avoids the considerable calculations of the FC layer, thereby improving the timeliness of feature learning. Moreover, it has greater flexibility in choosing the right network structure for more practical applications. In addition, the selection of medium-sized or large-sized networks can effectively reduce the need for feature pre-processing.

The structure and optimization process of the CNN-based feature learning are shown in Fig.4. We keep layers 1-7 (i.e., convolutional layers and FC layers) of our ImageNet pre-trained network (named pre-trained AlexNet) fixed, to initialize of the CNN-FC. CNN-FC network differs from AlexNet only in the classification layer. Then, we fine-tune the CNN-FC and replace the FC layers with KCRDP as a follow-up strategy. In the structure of AlexNet and CNN-FC, the 54.53 million parameters of the two FC layers create a large number of calculations, and the output 4096-dimensionality features are not conducive to low-dimensionality representation and efficient recognition, especially for recognition cases with fewer categories. When performing radar signal recognition, the TFI dataset is expressed as $\boldsymbol{S} \in R^{227\times227\times3\times N}$, where N is the total number of TFI samples. The l-dimensionality feature samples extracted from convolutional layers are expressed as $\boldsymbol{X} \in R^{l\times N}$, which send to the KCRDP layers to complete the classification and identification. The output features will be effectively reduced to m dimensionality ($m \ll l$).

B. THE KERNEL COLLABORATIVE REPRESENTATION AND DISCRIMINATIVE PROJECTION BASED CLASSIFICATION

Kernel mapping helps improve the feature space resolution. The discriminative projection in the kernel space reduces feature redundancy and enhances discriminative ability, and collaborative representation methods improve classification timeliness. In the l-dimensionality feature space, the training samples are expressed as $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_C] \in R^{l\times N}$, where $\boldsymbol{X}_i \in R^{l\times N_i}$ are the training samples of class i. The dictionary model considering $\tilde{l}$-dimensionality kernel space mapping and m-dimensionality discriminative space projection is

$$J_{(\boldsymbol{P},\boldsymbol{D},\boldsymbol{\alpha})} = \arg\min_{\boldsymbol{P},\boldsymbol{D},\boldsymbol{\alpha}} \left\| \boldsymbol{P}^{\mathrm{T}}\phi(\boldsymbol{y}) - \boldsymbol{P}^{\mathrm{T}}\phi(\boldsymbol{D})\boldsymbol{\alpha} \right\|_F^2 + \lambda \|\boldsymbol{\alpha}\|_F^2 \quad (3)$$

where $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_m] \in R^{\tilde{l}\times m}$ is the discriminative projection matrix, $\phi(\cdot) \in R^{\tilde{l}}$ is the kernel function, $\boldsymbol{\alpha} \in R^{\tilde{K}}$ is the coding coefficient, $\boldsymbol{P}$ and $\boldsymbol{D}$ can be represented linearly by using the atoms of the kernel sample space. Any column is represented as a linear combination of all kernel samples, such as $\boldsymbol{p}_j = \sum_{i=1}^{N} q_{j,i}\phi(\boldsymbol{x}_i)$. The discriminative projection matrix expressed as $\boldsymbol{P} = \phi(\boldsymbol{X})\boldsymbol{Q}$, where $\boldsymbol{Q} \in R^{N\times m}$ is the pseudo-random transformation matrix. The kernel dictionary is expressed as $\phi(\boldsymbol{D}) = \phi(\boldsymbol{X})\boldsymbol{A} = [\phi(\boldsymbol{D}_1), \phi(\boldsymbol{D}_2), \ldots, \phi(\boldsymbol{D}_C)]$, where $\phi(\boldsymbol{D}_i) = \phi(\boldsymbol{X}_i)\boldsymbol{A}_i \in R^{\tilde{m}\times\tilde{K}_i}$ is the sub-dictionary associated with each class i, and $\boldsymbol{A} \in R^{N\times\tilde{K}}$ is a dictionary coefficient matrix that satisfies

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & & & \\ & \boldsymbol{A}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_C \end{bmatrix} \quad (4)$$

where $\boldsymbol{A}_i \in R^{N_i\times\tilde{K}_i}$ is the sub-dictionary coefficient matrix corresponding to class i. Substituting the above representation, the dictionary learning model based on kernel collaborative representation and discriminant projection can be expressed as follows

$$J_{(\boldsymbol{Q},\boldsymbol{A},\boldsymbol{\alpha})} = \arg\min_{\boldsymbol{Q},\boldsymbol{A},\boldsymbol{\alpha}} \left\| \boldsymbol{Q}^{\mathrm{T}}K(\boldsymbol{X},\boldsymbol{y}) - \boldsymbol{Q}^{\mathrm{T}}K(\boldsymbol{X},\boldsymbol{X})\boldsymbol{A}\boldsymbol{\alpha} \right\|_F^2 + \lambda\|\boldsymbol{\alpha}\|_F^2 \quad (5)$$

where $K(\cdot)$ is the kernel gram function and $K(\boldsymbol{X},\boldsymbol{y}) = \phi(\boldsymbol{X})^{\mathrm{T}}\phi(\boldsymbol{y})$. $K(\boldsymbol{X},\boldsymbol{X}) = \phi(\boldsymbol{X})^{\mathrm{T}}\phi(\boldsymbol{X})$ is a semi-definite matrix. The coding coefficient of any sample $\boldsymbol{y}$ is given by

$$\hat{\boldsymbol{\alpha}} = \left(\boldsymbol{A}^{\mathrm{T}}K(\boldsymbol{X},\boldsymbol{X})^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}K(\boldsymbol{X},\boldsymbol{X})\boldsymbol{A} + \lambda I\right)^{-1} \cdot \boldsymbol{A}^{\mathrm{T}}K(\boldsymbol{X},\boldsymbol{X})^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}K(\boldsymbol{X},\boldsymbol{y}) \quad (6)$$

where $\hat{\boldsymbol{\alpha}} = [\hat{\boldsymbol{\alpha}}_1; \hat{\boldsymbol{\alpha}}_2; \cdots; \hat{\boldsymbol{\alpha}}_C] \in R^{\tilde{K}}$ is the coding coefficient, and $\hat{\boldsymbol{\alpha}}_i \in R^{\tilde{K}_i}$ is the coding coefficients belong to class i. The classification error of the sample associated with each dictionary i is defined as

$$e_i(\boldsymbol{y}) = \left\| \boldsymbol{Q}^{\mathrm{T}}K(\boldsymbol{X},\boldsymbol{y}) - \boldsymbol{Q}^{\mathrm{T}}K(\boldsymbol{X},\boldsymbol{X}_i)\boldsymbol{A}_i\hat{\boldsymbol{\alpha}}_i \right\|_F^2 \Big/ \left\|\hat{\boldsymbol{\alpha}}_i\right\|_F^2 \quad (7)$$

The category label is determined as $\mathrm{id}(\boldsymbol{y}) = \arg\min_i \{e_i\}$. Misclassification is mainly caused by differences between classes, which can be reduced by strengthening the differences between the correct type and the nearest misclassification. For any training sample $\boldsymbol{x}$, define $\tilde{E}^c$ as within-class

reconstruction error and $\tilde{E}^c = \left\| \boldsymbol{Q}^{\mathrm{T}} K(\boldsymbol{X}, \boldsymbol{x}) - \boldsymbol{Q}^{\mathrm{T}} K(\boldsymbol{X}, \boldsymbol{X}_c) \boldsymbol{A}_c \hat{\boldsymbol{\alpha}}_c \right\|_F^2$. Define $\tilde{E}^d$ as the minimum between-class reconstruction error, and the most adjacent misclassification class d can be expressed as

$$d = \arg\min_k \tilde{E}^k = \left\| \boldsymbol{Q}^{\mathrm{T}} K(\boldsymbol{X}, \boldsymbol{x}) - \boldsymbol{Q}^{\mathrm{T}} K(\boldsymbol{X}, \boldsymbol{X}_k) \boldsymbol{A}_k \hat{\boldsymbol{\alpha}}_k \right\|_F^2 \quad s.t.\ k \neq c \tag{8}$$

The batch representation and Fisher criterion [47], [49] are introduced to enhance the stability of the gradient update. The objective function of the error loss is defined as

$$J_B(\boldsymbol{Q}, \boldsymbol{A}) = \arg\min_{\boldsymbol{Q},\boldsymbol{A}} \frac{1}{B} \sum_{i=1}^{c} \sum_{j=1}^{B_i} \left(S\left(R\left(\boldsymbol{x}_{i,j}\right)\right) + \lambda \left\| \hat{\boldsymbol{a}}^{(i,j)} \right\|_F^2 \right) \tag{9}$$

where $S(\boldsymbol{x}) = (e^{\boldsymbol{x}} - e^{-\boldsymbol{x}})/(e^{\boldsymbol{x}} + e^{-\boldsymbol{x}})$ is the quantization function, $\boldsymbol{x}_{i,j}$ is the training sample of class i with number j, $B = \sum_{i=1}^{C} B_i$ is the batch size of the training set, B_i is the random sample size with each class i, and $\hat{\boldsymbol{\alpha}}^{(i,j)}$ is the coding coefficient solved by (6). The effectiveness and stability of gradient updates are guaranteed through batch random partitioning and batch iteration mechanisms. Define $R(\boldsymbol{x}) = \tilde{E}^c / \tilde{E}^d$ as the error quantization ratio. There are mainly two cases: (1) $\tilde{E}^c < \tilde{E}^d$, correct classification to class c can be realized, and the incorrect classification probability is reduced; (2)$\tilde{E}^c > \tilde{E}^d$, signals can be easily misclassified into class d. The MSGD algorithm [55] helps to correct the error.

During the optimization, partial derivatives of the pseudo-random transformation matrix and dictionary coefficient matrix are obtained by

$$g_{\boldsymbol{Q}} = \frac{\partial \boldsymbol{J}_B}{\partial \boldsymbol{Q}} = \frac{1}{B} \sum_i \sum_j S'\left(R\left(\boldsymbol{x}_{i,j}\right)\right) R\left(\boldsymbol{x}_{i,j}\right) \left(\frac{\nabla_{\boldsymbol{Q}} \tilde{E}^c_{i,j}}{\tilde{E}^c_{i,j}} - \frac{\nabla_{\boldsymbol{Q}} \tilde{E}^d_{i,j}}{\tilde{E}^d_{i,j}} \right)$$
$$g_{\boldsymbol{A}} = \frac{\partial \boldsymbol{J}_B}{\partial \boldsymbol{A}} = \frac{1}{B} \sum_i \sum_j S'\left(R\left(\boldsymbol{x}_{i,j}\right)\right) R\left(\boldsymbol{x}_{i,j}\right) \left(\frac{\nabla_{\boldsymbol{A}} \tilde{E}^c_{i,j}}{\tilde{E}^c_{i,j}} - \frac{\nabla_{\boldsymbol{A}} \tilde{E}^d_{i,j}}{\tilde{E}^d_{i,j}} \right) \tag{10}$$

where $S'(\boldsymbol{x}) = 1 - S^2(\boldsymbol{x})$ is the partial derivative of $S(\boldsymbol{x})$. Influences of $\hat{\boldsymbol{\alpha}}_g$ ($g \in \{c, d\}$) on the gradient are neglected, and the partial derivative of $\tilde{E}^g$ can be expressed as follows

$$\begin{aligned} \nabla_{\boldsymbol{Q}} \tilde{E}^g &= 2\left(K(\boldsymbol{X}, \boldsymbol{x}) - K\left(\boldsymbol{X}, \boldsymbol{X}_g\right) \boldsymbol{A}_g \hat{a}_g\right) \\ &\quad \cdot \left(\boldsymbol{Q}^{\mathrm{T}} K(\boldsymbol{X}, \boldsymbol{x}) - \boldsymbol{Q}^{\mathrm{T}} K\left(\boldsymbol{X}, \boldsymbol{X}_g\right) \boldsymbol{A}_g \hat{a}_g\right)^{\mathrm{T}} \\ \nabla_{\boldsymbol{A}} \tilde{E}^g &= -2\, K^{\mathrm{T}}\left(\boldsymbol{X}, \boldsymbol{X}_g\right) \boldsymbol{Q} \\ &\quad \cdot \left(\boldsymbol{Q}^{\mathrm{T}} K(\boldsymbol{X}, \boldsymbol{x}) - \boldsymbol{Q}^{\mathrm{T}} K(\boldsymbol{X}, \boldsymbol{X}) \boldsymbol{A}_g \hat{a}_g\right) \hat{a}_g^{\mathrm{T}} \end{aligned} \tag{11}$$

The gradient descent direction used to update $\boldsymbol{Q}$ and $\boldsymbol{A}$ under the MSGD algorithm can be expressed as follows

$$\begin{cases} \boldsymbol{Q}(t) = \boldsymbol{Q}(t-1) + \boldsymbol{U}(t) \\ \boldsymbol{A}(t) = \boldsymbol{A}(t-1) + \boldsymbol{V}(t) \end{cases} \tag{12}$$

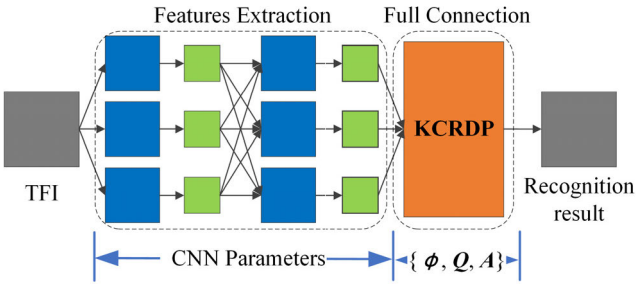

FIGURE 5. The CNN-KCRDP structure consists of a features extraction module and a full connection module.

where $\boldsymbol{U}(t) = \alpha_{\boldsymbol{Q}} \boldsymbol{U}(t-1) - \gamma g_{\boldsymbol{Q}}$ and $\boldsymbol{V}(t) = \alpha_{\boldsymbol{A}} \boldsymbol{V}(t-1) - \eta g_{\boldsymbol{A}}$. $\alpha_{\boldsymbol{Q}}$ and $\alpha_{\boldsymbol{A}}$ are the momentum factor, γ and η are learning rates. When the batch learning strategy is used, the gradient averaging is performed on (10). Hence, the learning rate varies with the batch parameters. Learning rates in case that $\gamma = B\gamma_0/C$ and $\eta = B\eta_0/C$, γ_0 and η_0 is the initial learning rate. The KCRDP optimization process based on MSGD is described in Algorithm 1.

Algorithm 1 The KCRDP Algorithm Based on MSGD

Input: Dictionary learning parameters of λ, m, N_1, and $\tilde{K}$; Kernel function; MSGD parameters of $\alpha_{\boldsymbol{Q}}$, $\alpha_{\boldsymbol{A}}$, γ_0, η_0, and B;

1: Perform samples kernel mapping and kernel matrix centralization. Initialize $\boldsymbol{A}^{(0)}$ as a Gaussian random matrix. Initialize $\boldsymbol{Q}^{(0)}$ by KPCA algorithm, $t = 0$;
2: **while** not converged **do**
3: $t \leftarrow t+1$;
4: Randomly divide the training set into *NUM* batches;
5: **for** $i = 1$: *NUM* **do**
6: Update the coding coefficient $\hat{\boldsymbol{\alpha}}^{(t)}$ by (6). Update $d^{(t)}$ by (8);
7: Update $g_Q^{(t)}$ and $g_A^{(t)}$ by (10) and (11);
8: Update $\boldsymbol{Q}^{(t)}$ and $\boldsymbol{A}^{(t)}$ by (12);
9: **end for**
10: **end while**

Output: $\boldsymbol{Q}$ and $\boldsymbol{A}$.

C. CNN-KCRDP BASED RECOGNITION

Fig.5 depicts the structure of CNN-KCRDP, in which the convolutional layers have the same structures as the corresponding classical CNN models, whereas all full connection layers are replaced with the KCRDP layer. As Fig.5 shows, the intercepted radar signal is converted into the time-frequency images. Then, the CNN extracted the multi-level nonlinear features by processing such as convolution and pooling. Learning appropriate low-dimensionality representations of deep features plays a critical role as a pre-processing procedure for the success of recognition, because of its convenience on geometric interpretations and its parsimony on computations [56]. The KCRDP is a classifier for DR learning

and discrimination. When ignoring the kernel mapping and collaborative representation, the parameters of KCRDP are approximately $N(m+\tilde{K})$, where m and $\tilde{K}$ are both relatively small, which significantly reduces the parameter quantity and feature dimensionality of the full connection layers. For radar signal modulation recognition, the CNN-KCRDP model is optimized by Algorithm 2. The recognition result is given by the smallest class obtained by formula (7).

Algorithm 2 The Deep Joint Learning Algorithm of CNN-KCRDP

Input: Radar intra-pulse signal and partial initialization parameters;

Initialization: Initialize CNN parameters by the AlexNet network, which is pre-trained through the ImageNet database;

Pro-processing: Perform formula (2) and bicubic interpolation algorithm to obtain the TFI dataset $\boldsymbol{S}$;

Optimization:

1: Update CNN parameters by fine-tuning the CNN-FC with a training set of TFI dataset;

2: Update KCRDP parameters by executing Algorithm 1;

Output: CNN-KCRDP parameters.

V. SIMULATION AND DISCUSSION

In this section, the effectiveness of the proposed technique is verified by experimental simulation, which mainly consists of five parts: First, the dataset of the radar signal and CNN-KCRDP's parameters are given. Second, feature stability is analysed. Then, the dictionary classification method is comprehensively compared with state-of-the-art methods. Moreover, the performance of the proposed technique is displayed and compared with other recognition algorithms. Finally, the robustness and timeliness of the algorithm are analysed. The specific simulation environment is built on a machine with CPU i7-9750H running at 2600 MHz, 6GB GPU (NVIDIA GeForce RTX 2060), 16GB RAM (DDR4@2666 MHZ), and Windows 10 (64bit). All generated data and recognition algorithms were simulated in MATLAB R2019b software, including pre-processing, and CNN-KCRDP optimization.

A. DATASET AND PARAMETERS SETTING

The simulations were carried out using the twelve radar signal modulation types given in section 3. Table 2 lists all parameters of the radar signal. $U(\cdot)$ is used to denote the uniform random distribution of data in the interval. $\{\cdot\}$ indicates a random parameter set. The f_s, L_c, f_c, N_{cc}, B, N, $f_{\min}$, M, ρ, and N_g represent the sampling frequency, length of barker codes, carrier frequency, number of carrier cycles in a single-phase symbol, bandwidth, number of samples, fundamental frequency, number of frequency steps, number of subcodes in a code, and number of segments, respectively. The phase states of the polytime codes are set to 2. The dataset was generated under an SNR that varied from -10dB to 8dB with a step of 2dB. Each type of signal produced 500 sets of data at each SNR, of which 80% was used for training, and 20% was used for testing.

TABLE 2. Radar signal parameters with different modulations.

Modulation type	Parameters	Ranges
	f_s	f_s =200MHz
BPSK	L_c	{7,11,13}
	f_c	$U(3f_s/20, f_s/5)$
	N_{cc}	{20,21,22,23}
LFM	f_c	$U(3f_s/20, f_s/5)$
	B	$U(f_s/20, f_s/10)$
	N	U[500,1000]
Costas	FH sequence	{3,4,5,6}
	$f_{\min}$	$U(f_s/48, f_s/30)$
	N	U[500,1000]
Frank, P1	f_c	$U(3f_s/20, f_s/5)$
	N_{cc}	{4,5,6}
	M	{5,6,7,8}
P2	f_c	$U(3f_s/20, f_s/5)$
	N_{cc}	{4,5,6}
	M	{6,8}
P3, P4	f_c	$U(3f_s/20, f_s/5)$
	N_{cc}	{4,5,6}
	ρ	{36,49,64}
T1, T2	f_c	$U(3f_s/20, f_s/5)$
	N_g	{4,5,6}
	N	U[500,1000]
T3, T4	f_c	$U(3f_s/20, f_s/5)$
	B	$U(f_s/20, f_s/10)$
	N	U[500,1000]

TABLE 3. Recognition performance of different kernel functions under -10dB environment.

Kernel function	RSR (%)
Gaussian	87.55
Polynomial	87.50
Linear	83.50

The Adam algorithm with a mini-batch size of 128 was used for fine-tuning the CNN. The maximum iteration and the initial learning rate were 50 and 0.001, respectively. The decay factor of first-order and second-order moment estimation was 0.9, 0.999, and the perturbation term was $1e^{-8}$. The Gaussian kernel function and regular parameter (λ) of 0.01 were used to optimize KCRDP, and the termination condition was that the maximum iterations reached 50 or the iteration error was less than 0.001. In the MSGD algorithm, momentum factors ($\alpha_{\boldsymbol{Q}}$ and $\alpha_{\boldsymbol{A}}$) were 0.5, and $\gamma_0 = 1$, $\eta_0 = 0.001$, and $B = 400$. The number of atoms was similar in each sub-dictionary, namely, $K_0 = 100$. The dimensionality reduction parameter m was 300.

Table 3 shows the recognition performance of different kernel functions under the -10dB environment, of which the Gaussian kernel function had better performance, and it was selected for subsequent processing.

We verify the influences of dictionary atoms and feature dimensionality on performance with numerous Monte Carlo simulations under the condition of SNR $= -10$dB. The results are shown in Fig.6. Comparative analysis indicates

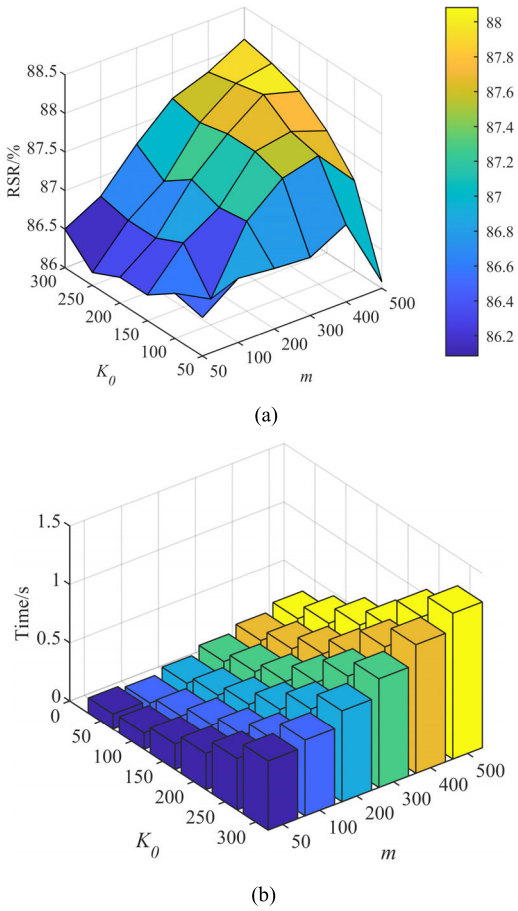

FIGURE 6. Influences of dictionary atoms and feature dimensionality at SNR of −10dB. (a) Influences on RSR. (b) Influences on testing time.

that the RSR and computational complexity are directly proportional to the dictionary atoms and feature dimensionality. The recognition performance is better when the dimensionality is higher than 300, and the number of atoms is higher than 100. Furthermore, the computational complexity mainly depends on the number of parameters included in $\boldsymbol{Q}$ and $\boldsymbol{A}$, which is positively correlated with the parameters m and K_0. Based on a comprehensive consideration of recognition performance and computational complexity, the relevant parameters are set to $m = 300$ and $K_0 = 100$.

B. FEATURE STABILITY ANALYSIS

Feature stability directly affects recognition performance. The recognition process of CNN-KCRDP includes four features: the CWD feature, pooling layer output features of pre-trained AlexNet, pooling layer output features of fine-tuned AlexNet, and low-dimensionality discriminative features of CNN-KCRDP. The 4800 samples of twelve radar signals are chosen under the condition of SNR = 8dB. Then, the extracted feature vector is reduced by the t-distributed stochastic neighbor embedding (TSNE) algorithm [29]. The result is shown in Fig.7. In the 2-D representation space of TSNE, the between-class difference of the original CWD feature and the pre-trained AlexNet feature is small, and there is a large degree of confusion. Moreover, the 9216-dimensionality fine-tuned AlexNet feature has a high degree of between-class discrimination and within-class aggregation. In addition, the 300-dimensionality discriminative feature of CNN-KCRDP maintains a considerable resolution equivalent to fine-tuned AlexNet feature, which is conducive to efficient subsequent identification, and indirectly demonstrates the feasibility of the joint learning model.

TABLE 4. Recognition results of different conventional algorithms and CNN-based classification algorithms.

	Algorithms	RSR (%)
Conventional algorithms	SVM	82.95
	KNN[8]	82.63
	KCRDP	84.75
	KCRC[33]	83.72
	LC-KSVD[34]	82.97
	JDDRDL[45]	83.97
	DSRC[47]	84.21
	KDL-DA[51]	84.36
CNN-based algorithms	CNN	96.22
	CNN+KCRDP	**97.80**
	CNN+JDDRDL	97.13
	CNN+DSRC	97.26
	CNN+KDL-DA	97.18
	CNN+RP+SVM	96.39
	CNN+PCA+SVM	96.72
	CNN+PCA+KNN	96.36
	CNN+PCA+KCRC	96.45
	CNN+PCA+LC-KSVD	95.02
	CNN+KPCA+SVM	96.89
	CNN+KPCA+KNN	96.55
	CNN+KPCA+KCRC	96.87
	CNN+KPCA+LC-KSVD	95.94

The performance of different deep network architectures, including AlexNet and CNN-KCRDP, is shown in Fig.8. Pre-trained AlexNet has inferior feature stability and generalization ability. Fine-tuning makes AlexNet more suitable for the radar signal dataset, which is conducive to forming a more stable representation. Moreover, the CNN-KCRDP performs better than fine-tuned AlexNet, especially in low SNR environments of −10-0dB.

C. DICTIONARY CLASSIFICATION COMPARISON

KCRDP is a dimensionality reduction and dictionary learning (DRDL) classifier, which projects the 9216-dimensionality output feature of CNN to a low-dimensionality discriminant space and performs efficient classification. To compare the advantages of KCRDP, it is compared with the classic DR algorithms [57], such as random projection (RP), principal component analysis (PCA), and kernel principal component analysis (KPCA). Additionally, it is compared with the traditional classifiers such as Gaussian support vector machines (SVM), k-nearest neighbor (KNN) [8], kernel collaborative representation classification (KCRC) [33], and LC-KSVD [14], [34]. The cross-validation method is used to select the optimal value of k of KNN. In addition, the performance of KCRDP is compared with several closely related

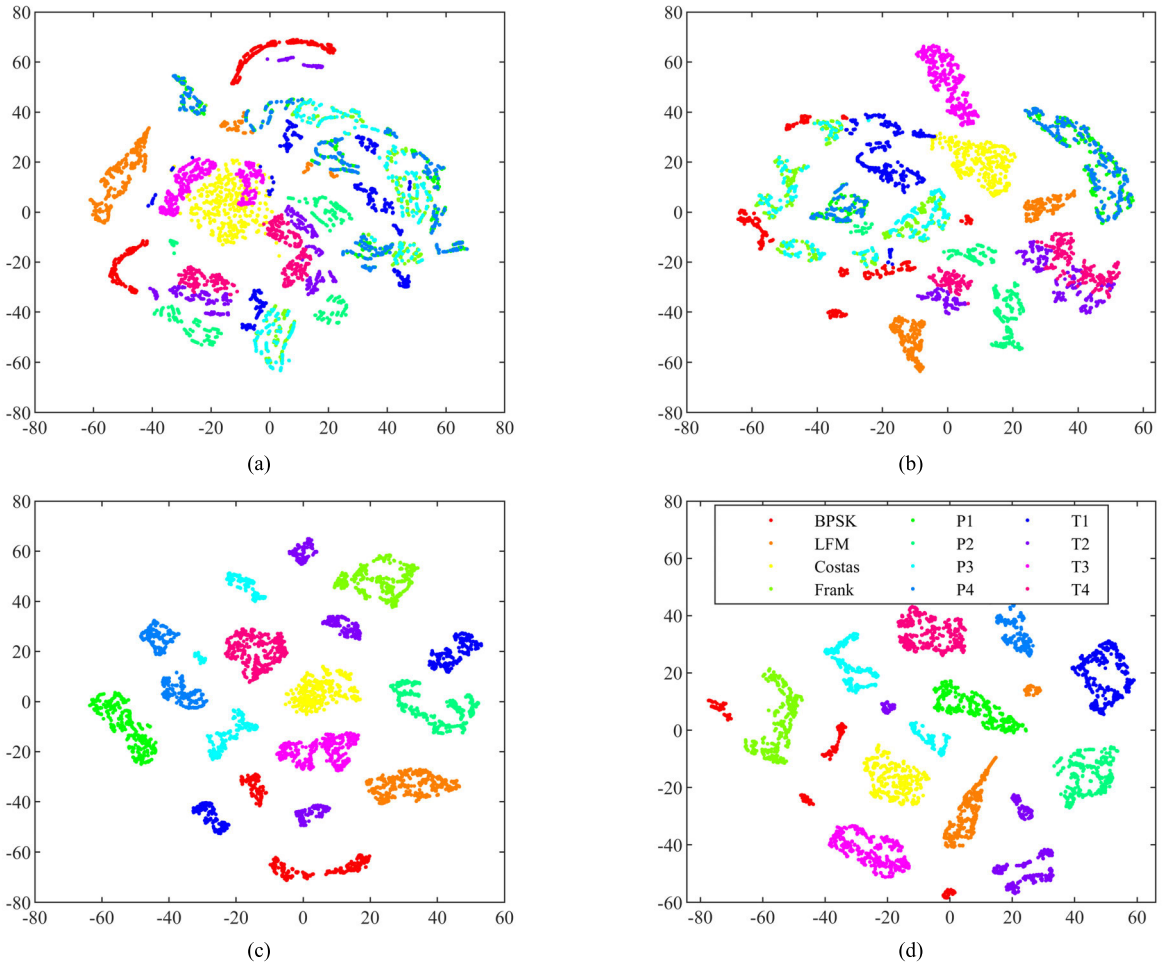

FIGURE 7. The 2-D feature map at SNR of 8dB. (a) CWD feature. (b) Pre-trained AlexNet feature. (c) Fine-tuned AlexNet feature. (d) Low-dimensionality discriminative feature of CNN-KCRDP.

JDDRDL [45], DSRC [47], and KDL-DA [51], which are dimensionality reduction and dictionary learning (DRDL) classification algorithms.

The overall average RSR under different classification algorithms are shown in Table 4, where CNN refers to the fine-tuned AlexNet. The recognition results of conventional algorithms are approximately 82.63% to 84.75%, of which the KCRDP is optimal. Moreover, they are all inferior to CNN-based methods, which achieved a higher RSR by more than 95.02%. CNN-based classification algorithms have better performance because the deep features extracted by the fine-tuned network are more stable than the CWD feature. In addition, the low-dimensionality features learned by the discriminative projection of DRDL algorithms are more targeted, and it is easier to form a stable representation than DR algorithms.

The computational complexity of these classification algorithms under a single SNR environment is shown in Table 5. Among them, machine learning methods have high timeliness of training; KCRC and LC-KSVD have better timeliness of training than DRDL methods, but the timeliness of testing is inferior to the JDDRDL and KCRDP methods, which improves the timeliness through discriminative projection and l_2 collaborative representation. It is worth noting that kernel mapping makes KCRDP more time-consuming than JDDRDL; l_1 regular constraint significantly reduces the timeliness of DSRC and KDL-DA. The results in Tables 4 and 5 demonstrate the superior of the KCRDP model in RSR and

TABLE 5. The computational complexity of different classification algorithms under single SNR.

Algorithms	Training time (s)	Testing time (s)
SVM	126.55	0.62
KNN	0.10	3.84
KCRDP	1141	0.29
KCRC	0	0.43
LC-KSVD	176.59	0.33
JDDRDL	1510.31	0.24
DSRC	912.27	401.16
KDL-DA	623.11	390.39

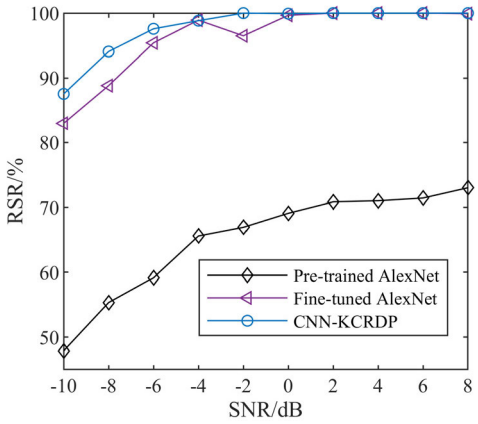

FIGURE 8. Performance comparison of different deep network architectures.

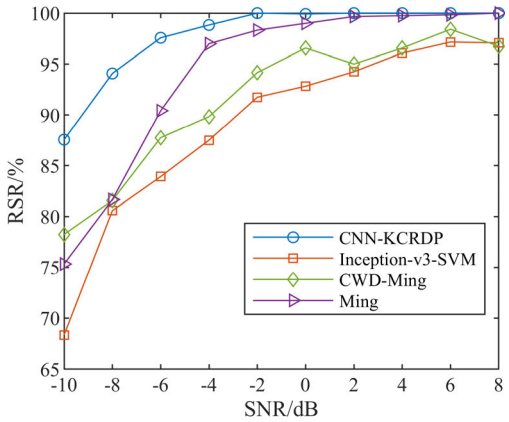

FIGURE 9. The comparison of different radar signal recognition methods.

test timeliness, but the optimization of KCRDP requires certain computing resources.

D. RECOGNITION METHOD COMPARISON

To further verify the effectiveness of the CNN-KCRDP method, the following recognition methods which have similar signal type or similar recognition architecture to this paper are selected for comparison: (1) the method of Ming [21]; (2) the Inception-v3-SVM method [29]; (3) the combination method of CWD feature and Ming's classification network which is denoted as CWD-Ming. The recognition results are shown in Fig.9 and Fig.10. As shown in Fig.9, the recognition performance under different recognition algorithms improves as the SNR increases. In addition, CNN-KCRDP has the best recognition performance, and the overall average RSR reaches 87.55% at −10dB. Moreover, the performance of the Inception-v3-SVM and CWD-Ming methods is relatively poor, it is difficult to identify multiple types of signals with high similarity effectively, and there is still a significant degree of confusion at high SNR.

The results show that CNN-KCRDP improves feature stability through a deep mapping mechanism, which can achieve optimal performance even without feature enhancement.

TABLE 6. Confusion matrix at SNR of −6dB.

Type	S1	S2	S3	S4	P1	P2	P3	P4	T1	T2	T3	T4
S1	100	0	0	0	0	0	0	0	0	0	0	0
S2	0	100	0	0	0	0	0	0	0	0	0	0
S3	0	0	100	0	0	0	0	0	0	0	0	0
S4	0	0	0	91	0	0	9	0	0	0	0	0
P1	0	0	0	0	92	0	0	8	0	0	0	0
P2	0	0	0	0	0	100	0	0	0	0	0	0
P3	0	0	0	6	0	0	94	0	0	0	0	0
P4	0	0	0	0	5	0	1	94	0	0	0	0
T1	0	0	0	0	0	0	0	0	100	0	0	0
T2	0	0	0	0	0	0	0	0	0	100	0	0
T3	0	0	0	0	0	0	0	0	0	0	100	0
T4	0	0	0	0	0	0	0	0	0	0	0	100

Its ability to recognize multiple types of high-similarity signals is significantly better than other methods, especially in the case of low SNR. Moreover, the recognition performance of Inception-v3-SVM is inferior to the other methods, because it is difficult to fine-tune and optimize the transfer learning network of Inception-v3 with small samples. Ming's classification network with only seven layers has difficulty forming an in-depth representation of features. The recognition result under the CWD feature is slightly better than Inception-v3-SVM. Furthermore, Ming's original paper used a feature enhancement pre-processing algorithm to enhance the time-frequency features, thereby reducing the capacity requirements of the classification network and performing better than the CWD-Ming.

Recognition performance of various signals with CNN-KCRDP under −6dB was selected for comparison. For convenience, BPSK, LFM, Costas, and Frank are abbreviated as S1 ∼ S4 in order. The confusion matrix is shown in Table 6. Specifically, Frank and P3, P1 and P4 have higher similarity and a relatively large degree of confusion. In addition, the overall average RSR reaches 97.58%.

E. ROBUSTNESS AND TIMELINESS ANALYSIS

To further explore the system robustness and timeliness, samples under SNR of −10dB, −2dB, and 6dB were selected for testing. There were 100 samples and 400 samples for each class under each SNR in the testing sets and training sets, respectively. The experiment was repeated fifty times, and 100-400 (with an interval of 100) samples were selected for testing in each training.

Fig.11 gives the relationship between the training size and the RSR under different SNR environments. As the training size increased, the overall average RSR gradually increased. The RSR of −10dB was approximately 77.33% when the training size was only 100, and it increased to 86.58% when the training size was 400. The RSR of −2dB and 6dB was always higher than 98.5% and gradually stabilized as the

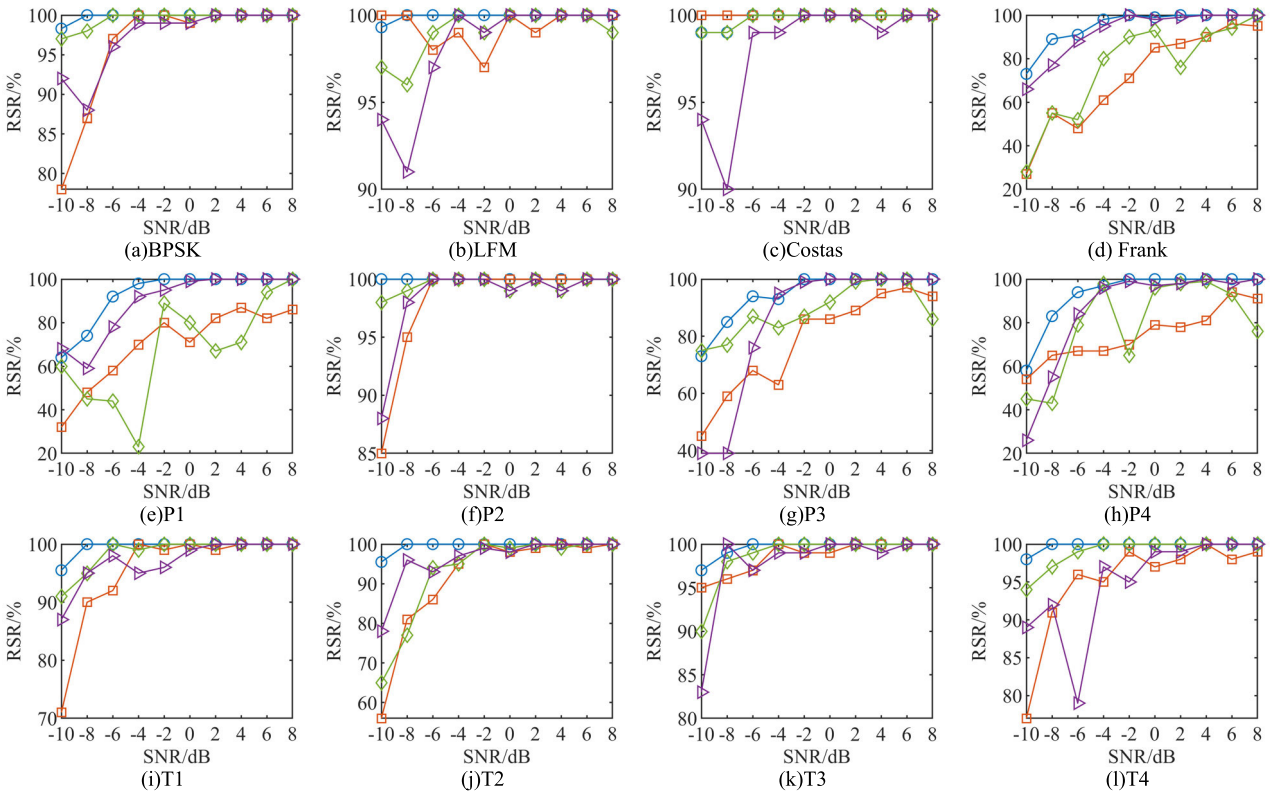

FIGURE 10. The comparison of signals with different modulations under different recognition methods.

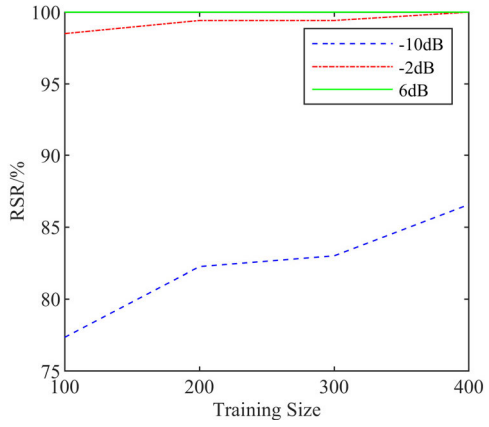

FIGURE 11. Recognition success rate under different training size at SNR of −10dB, −2dB, and 6dB.

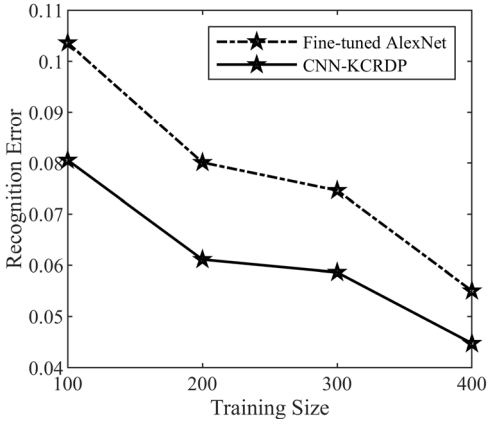

FIGURE 12. Recognition error of AlexNet and CNN-KCRDP under different training size.

training size increased. It shows strong adaptability to the small samples and low-SNR environments.

The recognition errors of AlexNet and CNN-KCRDP are shown in Fig.12. The recognition error decreased with the increase in the training size, and the difference between the AlexNet and the CNN-KCRDP gradually decreased with the increase in the training size, from 0.023 at 100 samples to 0.011 at 400 samples, which shows the advantages of this method over the AlexNet network on small samples.

The computational complexity test results under the same conditions are shown in Table 7. Both training and testing time are average running times under a single SNR. The training time of Inception-v3-SVM is mainly used to optimize the SVM classifier, which takes the shortest time. However, its deep features extraction process is very time-consuming because of the great depth of the Inception-v3 network. Ming's shallow network optimization takes less time, but the time consumption of the feature enhancement processing algorithm is considerable. The proposed CNN-KCRDP involves various optimizations and takes a long time, but the timeliness of the test is optimal because its convolutional

TABLE 7. Computational complexity comparison under single SNR.

Algorithms	Training time (s)	Testing time (s)
CNN-KCRDP	7142	79.19
Fine-tuned AlexNet	6001	128.83
Inception-v3-SVM[29]	35.87	114.36
Ming[21]	59.81	89.58

layers for feature extraction and KCRDP layers for classification have high timeliness. In addition, we use KCRDP layers instead of AlexNet's FC layers, which reduce network parameters and have better test timeliness than AlexNet. In general, supervised learning is completed through offline learning. Time consumption during training can be neglected temporarily. System timeliness mainly depends on testing time. As shown by the overall comparison results of RSR and computational complexity, the proposed method has certain advantages in recognition accuracy and recognition efficiency.

Combining with the simulation results, it is concluded that for the recognition process shown in Fig.2, CNN mainly affects the upper limit of recognition performance, and its deep mapping greatly promotes the improvement of RSR. Compared with dictionary learning methods, CNN-KCRDP architecture has multi-level representation capabilities and better recognition performance. Compared with deep networks, CNN-KCRDP is more suitable for small samples and more time-saving for classification, and it more easily forms low-dimensionality representations. Therefore, the combination of deep learning networks and dictionary classification can effectively improve feature stability, and further improve timeliness and classification performance.

In practical applications, the application of deep networks is limited by conditions such as sample capacity and computing resources. Deep networks reduce the need for feature enhancement processing and have strong recognition performance with the support of large samples and computing resources. The advantage of small networks (such as Ming's classification network) is high timeliness, but they are not well suited for classification tasks in low-SNR environments and multiple signal types due to insufficient depth representation capabilities. In addition, small networks have a greater need for feature enhancement processing. For applications with relatively low complexity such as radar signal identification, medium-sized networks shown in this paper can be reasonably selected for efficient classification and identification, and it has performance advantages when the sample set is relatively small. For more complicated applications, we can choose a deeper network to replace the convolutional layers of CNN-KCRDP.

VI. CONCLUSION

In this paper, we propose a new multi-level representation-based deep joint learning architecture for modulation recognition, which combines the advantages of deep learning, kernel collaborative representation, and discriminative projection. The recognition system is built on the CWD feature without feature enhancement processing. Deep features are obtained through the fine-tuned AlexNet's convolutional layers, which enhances the feature nonlinearity and stability. Furthermore, the discriminative projection and classification of low-dimensionality kernel space are realized through KCRDP, which improves the classification ability of small samples and the timeliness. Moreover, the feature stability, dictionary classification performance and deep joint learning technology performance are verified under the radar signal dataset. The experimental results under twelve different modulation signals show stronger competitive performance in comparison with the AlexNet network and other recognition algorithms, in terms of timeliness and adaptability to small samples. The overall average RSR aiming at twelve radar signal modulation types reached 97.58% at SNR of −6dB. In addition, the deep joint learning framework can also be applied to more recognition tasks or dimensionality reduction learning tasks.

REFERENCES

[1] V. Iglesias, J. Grajal, P. Royer, M. A. Sanchez, M. Lopez-Vallejo, and O. A. Yeste-ojeda, "Real-time low-complexity automatic modulation classifier for pulsed radar signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 108–126, Jan. 2015.
[2] G. A. Montazer, H. Khoshniat, and V. Fathi, "Improvement of RBF neural networks using fuzzy-OSD algorithm in an online radar pulse classification system," *Appl. Soft Comput.*, vol. 13, no. 9, pp. 3831–3838, Sep. 2013.
[3] Q. Guo, P. Nan, and J. Wan, "Radar signal recognition based on ambiguity function features and cloud model similarity," in *Proc. 8th Int. Conf. Ultrawideband Ultrashort Impulse Signals (UWBUSIS)*, Odesa, Ukraine, Sep. 2016, pp. 128–134.
[4] Z. Zhang, Y. Li, S. Jin, Z. Zhang, H. Wang, L. Qi, and R. Zhou, "Modulation signal recognition based on information entropy and ensemble learning," *Entropy*, vol. 20, no. 3, p. 198, Mar. 2018.
[5] S. B. S. Hanbali and R. Kastantin, "Automatic modulation classification of LFM and polyphase-coded radar signals," *Radioengineering*, vol. 26, no. 4, pp. 1118–1125, Dec. 2017.
[6] Z. Yang, Z. Wu, Z. Yin, T. Quan, and H. Sun, "Hybrid radar emitter recognition based on rough K-means classifier and relevance vector machine," *Sensors*, vol. 13, no. 1, pp. 848–864, Jan. 2013.
[7] X. Ma, D. Liu, and Y. Shan, "Intra-pulse modulation recognition using short-time Ramanujan Fourier transform spectrogram," *EURASIP J. Adv. Signal Process.*, vol. 1, Jun. 2017, Art. no. 42.
[8] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2742–2750, Aug. 2012.
[9] L. T. Liu, S. Wang, and Z. K. Zhao, "Radar waveform recognition based on time-frequency analysis and artificial bee colony-support vector machine," *Electronics*, vol. 7, no. 5, p. 59, Apr. 2018.
[10] M. Zhang, L. Liu, and M. Diao, "LPI radar waveform recognition based on time-frequency distribution," *Sensors*, vol. 16, no. 10, p. 1682, Oct. 2016.
[11] R. Cao, J. Cao, J.-P. Mei, C. Yin, and X. Huang, "Radar emitter identification with bispectrum and hierarchical extreme learning machine," *Multimedia Tools Appl.*, vol. 78, no. 20, pp. 28953–28970, Oct. 2019.
[12] S. Tariyal, A. Majumdar, R. Singh, and M. Vatsa, "Deep dictionary learning," *IEEE Access*, vol. 4, pp. 10096–10109, 2016.
[13] C. Bao, H. Ji, Y. Quan, and Z. Shen, "Dictionary learning for sparse coding: Algorithms and convergence analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1356–1369, Jul. 2016.
[14] H. Wang, M. Diao, and L. Gao, "Low probability of intercept radar waveform recognition based on dictionary learning," in *Proc. 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Hangzhou, China, Oct. 2018, pp. 1–6.

[15] Z. Zhou, G. Huang, H. Chen, and J. Gao, "Automatic radar waveform recognition based on deep convolutional denoising auto-encoders," *Circuits, Syst., Signal Process.*, vol. 37, no. 9, pp. 4034–4048, Sep. 2018.
[16] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, and J. T. Zhou, "Deep clustering with sample-assignment invariance prior," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.
[17] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5076–5086, Oct. 2018.
[18] P. Hu, D. Peng, Y. Sang, and Y. Xiang, "Multi-view linear discriminant analysis network," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5352–5365, Nov. 2019.
[19] P. Hu, D. Peng, X. Wang, and Y. Xiang, "Multimodal adversarial network for cross-modal retrieval," *Knowl.-Based Syst.*, vol. 180, pp. 38–50, Sep. 2019.
[20] X. Wang, G. Huang, Z. Zhou, W. Tian, J. Yao, and J. Gao, "Radar emitter recognition based on the energy cumulant of short time Fourier transform and reinforced deep belief network," *Sensors*, vol. 18, no. 9, p. 3103, Sep. 2018.
[21] M. Zhang, M. Diao, and L. Guo, "Convolutional neural networks for automatic cognitive radio waveform recognition," *IEEE Access*, vol. 5, pp. 11074–11082, 2017.
[22] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 2437–2441.
[23] L. Gao, X. Zhang, J. Gao, and S. You, "Fusion image based radar signal feature extraction and modulation recognition," *IEEE Access*, vol. 7, pp. 13135–13148, 2019.
[24] M. Kong, J. Zhang, W. Liu, and G. Zhang, "Radar emitter identification based on deep convolutional neural network," in *Proc. Int. Conf. Control, Autom. Inf. Sci. (ICCAIS)*, Hangzhou, China, Oct. 2018, pp. 309–314.
[25] M. Zhang, M. Diao, L. Gao, and L. Liu, "Neural networks for radar waveform recognition," *Symmetry*, vol. 9, no. 5, p. 75, May 2017.
[26] S.-H. Kong, M. Kim, L. M. Hoang, and E. Kim, "Automatic LPI radar waveform recognition using CNN," *IEEE Access*, vol. 6, pp. 4207–4219, 2018.
[27] Z. Qu, X. Mao, and Z. Deng, "Radar signal intra-pulse modulation recognition based on convolutional neural network," *IEEE Access*, vol. 6, pp. 43874–43884, 2018.
[28] Z. Qu, W. Wang, C. Hou, and C. Hou, "Radar signal intra-pulse modulation recognition based on convolutional denoising autoencoder and deep convolutional neural network," *IEEE Access*, vol. 7, pp. 112339–112347, 2019.
[29] Q. Guo, X. Yu, and G. Ruan, "LPI radar waveform recognition based on deep convolutional neural network transfer learning," *Symmetry*, vol. 11, no. 4, p. 540, Apr. 2019.
[30] J. Wan, X. Yu, and Q. Guo, "LPI radar waveform recognition based on CNN and TPOT," *Symmetry*, vol. 11, no. 5, p. 725, May 2019.
[31] C.-Y. Lee, S. Xie, Z. Zhang, Z. Tu, and P. Gallagher, "Deeply-supervised nets," in *Proc. 18th Int. Conf. Artif. Intell. Statist.*, 2015, pp. 562–570.
[32] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 471–478.
[33] W. Yang, Z. Wang, J. Yin, C. Sun, and K. Ricanek, "Image classification using Kernel collaborative representation with regularized least square," *Appl. Math. Comput.*, vol. 222, pp. 13–28, Oct. 2013.
[34] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.
[35] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5160–5175, Nov. 2017.
[36] J. Sulam, V. Papyan, M. Elad, and Y. Romano, "Multilayer convolutional sparse modeling: Pursuit and dictionary learning," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4090–4104, Aug. 2018.
[37] Z. Zhang, W. Jiang, Z. Zhang, S. Li, G. Liu, and J. Qin, "Scalable block-diagonal locality-constrained projective dictionary learning," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Macao, China, Aug. 2019, pp. 4376–4382.
[38] Z. Zhang, J. Ren, W. Jiang, Z. Zhang, R. Hong, S. Yan, and M. Wang, "Joint subspace recovery and enhanced locality driven robust flexible discriminative dictionary learning," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
[39] S. Gu, L. Zhang, X. Feng, and W. Zuo, "Projective dictionary pair learning for pattern classification," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, Dec. 2014, pp. 793–801.
[40] Z. Zhang, W. Jiang, J. Qin, L. Zhang, F. Li, M. Zhang, and S. Yan, "Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3798–3814, Aug. 2018.
[41] Z. Zhang, Y. Sun, Z. Zhang, Y. Wang, G. Liu, and M. Wang, "Learning structured twin-incoherent twin-projective latent dictionary pairs for classification," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Beijing, China, Nov. 2019, pp. 836–845.
[42] Z. Kang, H. Pan, S. C. H. Hoi, and Z. Xu, "Robust graph learning from noisy data," *IEEE Trans. Cybern.*, to be published.
[43] J. Ren, Z. Zhang, S. Li, Y. Wang, G. Liu, S. Yan, and M. Wang, "Learning hybrid representation by robust dictionary learning in factorized compressed space," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 3941–3956, Jan. 2020.
[44] Y. Sun, Z. Zhang, W. Jiang, Z. Zhang, L. Zhang, S. Yan, and M. Wang, "Discriminative local sparse representation by robust adaptive dictionary pair learning," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.
[45] Z. Feng, M. Yang, L. Zhang, Y. Liu, and D. Zhang, "Joint discriminative dimensionality reduction and dictionary learning for face recognition," *Pattern Recognit.*, vol. 46, no. 8, pp. 2134–2143, Aug. 2013.
[46] B.-Q. Yang, C.-C. Gu, K.-J. Wu, T. Zhang, and X.-P. Guan, "Simultaneous dimensionality reduction and dictionary learning for sparse representation based classification," *Multimedia Tools Appl.*, vol. 76, no. 6, pp. 8969–8990, Mar. 2017.
[47] H. Zhang, Y. Zhang, and T. S. Huang, "Simultaneous discriminative projection and dictionary learning for sparse representation based classification," *Pattern Recognit.*, vol. 46, no. 1, pp. 346–354, Jan. 2013.
[48] Y. Qin, X. Bian, and Y. Sheng, "Multiple Kernel learning for representation-based classification of hyperspectral images," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2018, pp. 3507–3512.
[49] G. Zhang, H. Sun, G. Xia, and Q. Sun, "Kernel collaborative representation based dictionary learning and discriminative projection," *Neurocomputing*, vol. 207, pp. 300–309, Sep. 2016.
[50] Y. Chi and F. Porikli, "Classification and boosting with multiple collaborative representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1519–1531, Aug. 2014.
[51] G. Zhang, H. Sun, Z. Ji, G. Xia, L. Feng, and Q. Sun, "Kernel dictionary learning based discriminant analysis," *J. Vis. Commun. Image Represent.*, vol. 40, pp. 470–484, Oct. 2016.
[52] E. L. Key, "Detecting and classifying low probability of intercept radar," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 6, pp. 39–41, Jun. 2004.
[53] J. W. Hwang and H. S. Lee, "Adaptive image interpolation based on local gradient features," *IEEE Signal Process. Lett.*, vol. 11, no. 3, pp. 359–362, Mar. 2004.
[54] U. Cote-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.
[55] H. Li, K. Li, J. An, and K. Li, "MSGD: A novel matrix factorization approach for large-scale collaborative filtering recommender systems on GPUs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 29, no. 7, pp. 1530–1544, Jul. 2018.
[56] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[57] H. Binol, "Ensemble learning based multiple Kernel principal component analysis for dimensionality reduction and classification of hyperspectral imagery," *Math. Problems Eng.*, vol. 2018, pp. 1–14, Sep. 2018.

DONGJIN LI was born in Guangyuan, China, in 1992. He received the B.S. and M.S. degrees from Air Force Early Warning Academy, Wuhan, China, in 2014 and 2017, respectively, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include deep learning, dictionary learning, radar signal processing, machine learning, and intelligent application of integrated systems.

RUIJUAN YANG was born in Zhongjiang, China, in 1964. She received the B.S. and M.S. degrees from the Chengdu University of Electronic Science and Technology, China, in 1984 and 1987, respectively, and the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China.

She is currently a Professor with Air Force Early Warning Academy, Wuhan. She is a high-level air force talent and an air force-level expert. She has completed more than 20 scientific major research projects both inside and outside the military. She has authored more than 60 articles. She holds four patents. She has been granted one software copyright. Her research interests include communication technology, radar signal processing, integrated radar communication technology, machine learning, and intelligent application of integrated systems. She received seven military science and technology talent progress awards.

XIAOBAI LI was born in Longxi, China, in 1983. He received the B.S., M.S., and Ph.D. degrees from Radar Academy, Wuhan, China, in 2006, 2009, and 2013, respectively. He is currently a Lecturer with the Laboratory of Radar and Communication Integrated Technology, Air Force Early Warning Academy. His current research interests include integrated radar communication technology, machine learning, radar signal processing, and waveform design and optimization.

SHENGKUN ZHU was born in Yinan, China, in 1996. He received the B.S. degree in information and communication engineering from the Dalian University of Technology, Dalian, China, in 2018. He is currently pursuing the M.S. degree with Air Force Early Warning Academy, Wuhan, China. His current research interests include radar signal processing, radar waveform design, and intelligent application of integrated systems.