Multi-kernel and Multi-task Learning for Radar Target Recognition

Cong Li, Xianyu Wang(✉), and Xu Yang

Academy of Space Electronic Information Technology, Xi'an 710100, China
lcongxd@126.com

Abstract. In this paper, a multiple kernel and multiple task learning framework (MKMTL) is proposed. To improve the interpretability of input data and adapt to different data sets, a weighted data-dependent kernel function is proposed and extended to multiple kernel functions. To fully reveal and utilize the shared information among different radar targets, multi-task learning framework is proposed. In this paper, a larger class of mixed norm penalty is adopted. It can increase the flexibility of MKMTL model. To verify the performance of the proposed model, measured MSTAR SAR public database is conducted. Experimental results demonstrate that the proposed method can effectively utilize the shared or potential information among different tasks and exhibits a better recognition performance compared with several popular existing recognition methods.

Keywords: Multi-kernel learning (MKL) · Multi-task learning (MTL) · Radar target recognition · Synthetic aperture radar (SAR)

1 Introduction

Based on radar images, automatic target recognition has the ability to produce 24-hour-a-day and robustness towards all-weather condition, which aroused widespread concern among researchers. For the past few years, SAR images have been extensively studied in ATR fields [1–5]. One of the biggest challenges for radar target recognition is that SAR images are highly sensitive to the variation of pose and speckle noise. We still need to study how to recognize the specified radar target. In radar ATR three main stages (detection, discrimination, and classification) are included. Determining the presence and location of the target is completed in the first phase. Then the background clutter is suppressed in the second stage. In the last stage, the category of target is determined by the designed classifier. In this paper, we emphasize features discrimination and classifier design.

Different from the features of optical target, the geometric structure features of radar target hidden in the radar echo are considerably nonlinear and complicated. In order to effectively reveal the nonlinear features of radar targets, the kernel trick is widely utilized [6–10]. By the kernel trick, the input space is mapped into a high-dimensional space, where the linearly indivisible features become separable in the high dimension space. Nevertheless, most of these kernel functions are monotonous and inflexible, which restricts their performances. To adapt to different data sets, a more flexible

B. Li et al. (Eds.): IoTaaS 2020, LNICST 346, pp. 380–391, 2021.
https://doi.org/10.1007/978-3-030-67514-1_31

learning method, called multiple kernel learning (MKL), is proposed [8, 11–13]. Wang et al. propose a discriminative multiple kernel learning (DMKL) method for spectral image classification, where an optimal combined kernel is learned by maximizing separability in reproduction kernel Hilbert space with Fisher criterion (FC) and maximum margin criterion (MMC) [11]. Gu et al. adopt low-rank nonnegative matrix factorization (NMF) and kernel NMF (KNMF) to enhance the ability of unsupervised learning with the predefined base kernels [12]. Besides, Gu et al. propose a multiple kernel sparse representation classification (MKSRC) framework for land cover classification, in which multiple kernel learning (MKL) is embedded into sparse representation classification (SRC) [13]. All the experimental results demonstrate that multiple kernel model is more flexible and can achieve a better recognition performance compared with single kernel. Based on the advantages of multi-kernel learning in ATR field, a weighted data-dependent multiple kernel function is proposed in this paper.

Although multi-kernel learning methods perform well in the field of target recognition, most of these methods are implemented within the framework of single task, which ignores the latent relatedness among multiple tasks. To make full use of the shared information among multi-tasks, a multi-task learning (MTL) framework is proposed [1, 14–19]. Zheng et al. propose a multi-task learning to decrease the within-class distance and increase the between-class scatter for face and expression recognition [18]. Liu et al. propose a method for cells classification by clustered multi-task learning [1]. Dong et al. propose a multi-task learning model for SAR target recognition, where each component of monogenic signal is specified as one single task [19]. These works have provided empirical evidence on the benefit of multi-task framework. Most of the multi-task relationships of these methods are realized by the fixed regularization principle $\ell 1 - \ell 2$. To further increase the flexibility of multi-task model, a larger class of mixed norm penalty $\ell_p - \ell_q$ [20, 21] is applied to this paper.

Inspired by the advantage of multi-kernel and multi-task learning, a multiple kernel and multiple task learning framework (MKMTL) is proposed in this paper and three parts of work have been done. To improve the flexible of kernel function, a data-dependent kernel function [22] is selected as the basic kernel. Based on the basic kernel, the weighted kernel is generated according to Kernel Alignment (KA) measure criterion [23]. Then, the weighed kernel is extended to multikernel by varying the bandwidth parameters. The realization of multi-task learning is the second part of work. In this paper, each of the target recognition or multi-kernel learning is considered as a single task and different tasks to be learned share a common subset of kernel representations. In this paper, a more flexibly mixed norm penalty $\ell_p - \ell_q$ is utilized, where ℓ_p controls the sparsity of the kernel representations across tasks and ℓ_q determines the importance of the task relatedness. To quickly obtain the optimal combinations of model parameters, a genetic algorithm is utilized, which is the last part of work. Lastly, comprehensive experiments on simulated HRRP dataset and measured MSTAR SAR public database are conducted to verify effectiveness of the proposed method.

The rest of the paper is organized as follows. The data-dependent multi-kernel and multi-task learning (MKMTL), and the MKMTL framework for radar target recognition are proposed in Sect. 2. Then the simulation experiments and results analysis based on HRRP and SAR data sets are given in Sect. 3. Finally, conclusions are drawn in Sect. 4.

2 Multi-kernel and Multi-task Learning

2.1 Multiple Kernel Learning (MKL) Based on Data-Dependent Kernel Function

Suppose we are given N samples $\{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in R^n$ is input vector with the target output $y_i \in \{1, -1\}$. It has been proved that multiple kernels have a better interpretability of the decision function and can improve the performances [8, 11–13]. A convenient and efficient approach is to consider a convex combination of basis kernels

$$K(x_i, x) = \sum_{m=1}^{M} d_m K_m(x_i, x), \tag{1}$$

with $d_m \geq 0$ and $\sum_{m=1}^{M} d_m = 1$. In Eq. (1), $K_m(x_i, x)$ is the basic kernel, d_m is the coefficient of basic kernel function, and M is the total number of kernels. In this paper, a data-dependent kernel is used as the basic kernel K_m and the corresponding weight is defined as the coefficient d_m. To be specific, the data-dependent kernel function [22] can be expressed as

$$K(x_i, x_j) = q(x_i) q(x_j) \tilde{K}(x_i, x_j), \tag{2}$$

where $\tilde{K}(x_i, x_j)$ is defined as the Gaussian kernel. $q(x) = \beta_0 + \sum_{i=1}^{k} \beta_i \hat{K}(x, \widehat{x}_i)$ is the factor function, and $\hat{K}(x, \widehat{x}_i) = \exp\left(-\gamma \left\| x - \widehat{x}_i \right\|^2\right)$, β_i is the combination coefficient, $\{\widehat{x}_i \in R^n, i = 1, \ldots, k\}$ is called the "empirical cores". $q(x)$ is determined by the data itself and can effectively describe the characteristics of different data. Therefore, Eq. (2) is used as the data-dependent kernel. Suppose the kernel matrices corresponding to $\tilde{K}(x_i, x_j)$ and $K(x_i, x_j)$ are denoted by $\tilde{K}$ and K respectively. Then K can be easily expressed as $K = Q\tilde{K}Q$, where Q is a diagonal matrix, whose diagonal elements are $\{q(x_1), q(x_2), \ldots, q(x_N)\}$. Let $q = (q(x_1), q(x_2), \ldots, q(x_N))^T$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_K)^T$, then q can be rewritten as

$$q = \begin{pmatrix} 1 & \hat{K}(x_1, \widehat{x}_1) & \ldots & \hat{K}(x_1, \widehat{x}_K) \\ 1 & \hat{K}(x_2, \widehat{x}_1) & \ldots & \hat{K}(x_2, \widehat{x}_K) \\ \vdots & & & \\ 1 & \hat{K}(x_N, \widehat{x}_1) & \ldots & \hat{K}(x_N, \widehat{x}_K) \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} = \bar{K}\beta. \tag{3}$$

To evaluate the classification performance of kernel function, the separable measure between two classes is taken as the criterion

$$J = \frac{q^T \tilde{B} q}{q^T \tilde{W} q}, \tag{4}$$

where $\tilde{B}$ and $\tilde{W}$ are the between-class and the within class kernel scatter matrices of basic kernel $\tilde{K}$ respectively. Suppose L tasks are to be learned, then $\tilde{B}$ and $\tilde{W}$ can be denoted respectively as

$$\tilde{B} = \begin{pmatrix} \frac{1}{N_1}K_{11} & 0 & \cdots & 0 \\ 0 & \frac{1}{N_2}K_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{N_L}K_{LL} \end{pmatrix} - \frac{1}{N}\tilde{K} \tag{5}$$

$$\tilde{W} = \begin{pmatrix} k_{11} & 0 & \ldots & 0 \\ 0 & k_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & k_{NN} \end{pmatrix} - \begin{pmatrix} \frac{1}{N_1}K_{11} & 0 & \cdots & 0 \\ 0 & \frac{1}{N_2}K_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{N_L}K_{LL} \end{pmatrix}, \tag{6}$$

where $K_{ii} \in R^{N_i} \times R^{N_i}$, $(i = 1, 2, \ldots, L)$ is the kernel matrices of $i - th$ class samples. The coefficient β can be obtained by using the standard gradient descent method

$$\beta^{(n+1)} = \beta^{(n)} + \eta\left(\frac{\bar{K}_1^T \tilde{B} \bar{K}}{(q^{(n)})^T \tilde{W} q^{(n)}} - J(\beta^{(n)}) \frac{\bar{K}_1^T \tilde{W} \bar{K}}{(q^{(n)})^T \tilde{W} q^{(n)}}\right)\beta^{(n)}, \tag{7}$$

where η is a learning rate. Based on the obtained data-dependent kernel $K(x_i, x_j)$, a set of kernels can be easily generated by varying the width parameter of the basic Gaussian kernel function.

To further improve the class separability of multiple kernel functions, the weighted multiple kernels are generated based on the Kernel Alignment (KA) measure [23], where the similarity between the input kernel and ideal kernel are measured. The alignment between the data-dependent kernel $\boldsymbol{K}_{data}$ and the ideal kernel $\boldsymbol{K}_{ideal}$ can be written as

$$KA(\boldsymbol{K}_{data}, \boldsymbol{K}_{ideal}) = \frac{\langle \boldsymbol{K}_{data}, \boldsymbol{y}\boldsymbol{y}^T \rangle_F}{\sqrt{\langle \boldsymbol{K}_{data}, \boldsymbol{K}_{data} \rangle_F \langle \boldsymbol{y}\boldsymbol{y}^T, \boldsymbol{y}\boldsymbol{y}^T \rangle_F}}, \tag{8}$$

where $\langle .,. \rangle_F$ stands for the Frobenius distance $\langle \boldsymbol{K}_x, \boldsymbol{K}_y \rangle_F = \sum_{i,j} \boldsymbol{K}_x(x_i, x_j) \boldsymbol{K}_y(x_i, x_j)$. Then, a proportional weighting is applied as follows

$$\zeta_m = \frac{KA(\boldsymbol{K}_m, \boldsymbol{K}_{ideal})}{\sum_{m=1}^{M} KA(\boldsymbol{K}_m, \boldsymbol{K}_{ideal})} \quad \forall m. \tag{9}$$

The weighted data-dependent kernel function can be written as

$$K_m(x_i, x_j) = \zeta_m k_{data} = \zeta_m q(x_i) q(x_j) \tilde{K}_m(x_i, x_j). \tag{10}$$

where $\zeta_m = d_m$ is the coefficient of basic kernel function. The objective function of MKL based on data-dependent kernel function can be formulated as

$$\begin{cases} \min\limits_{d} \max\limits_{\alpha} \sum\limits_{i=1}^{N} \alpha_i - \frac{1}{2} \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \alpha_i \alpha_j y_i y_j \sum\limits_{m=1}^{M} d_m K_m(x_i, x_j) \\ s.t. \quad \sum\limits_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, \quad i = 1, 2 \ldots N \end{cases}. \tag{11}$$

Equation (11) shows that MKL aims to learn the Lagrange multipliers parameters α and the best weight vector d for the linear combination of base kernels.

2.2 Multiple Task Learning (MTL) Based on Mixed Norm Penalty $\ell_p - \ell_q$

Suppose L classifiers (tasks) are to be trained (learned) from L different datasets $\{(x_{i,1} y_{i,1})_{i=1}^{N_1}, (x_{i,2} y_{i,2})_{i=1}^{N_2}, \ldots, (x_{i,L} y_{i,L})_{i=1}^{N_L}\}$ with $\sum\limits_{t=1}^{L} N_t = N$. For the t th task, the decision function is give as

$$f_t(x) = \sum_{i=1}^{N_t} \sum_{m=1}^{M} \alpha_{t,i} d_{t,m} K_m(x_i, x) + b_t. \tag{12}$$

In order to be more data-adaptive, a larger class of mixed norm penalty $\ell_p - \ell_q$ [20, 21] is applied to the regularization term

$$\Omega_{p,q}(f_1, \ldots, f_L)^2 = \left(\sum_{m=1}^{M} \left(\sum_{t=1}^{L} \|f_{t,m}\|_{\mathrm{H}_m}^{q}\right)^{\frac{p}{q}}\right)^2, \tag{13}$$

where ℓ_p controls the sparsity of the kernel representations across tasks and ℓ_q determines the importance of the task relatedness. In the simulation experiments, it was found that $\ell_p (p = 1)$ will produce the desired model sparsity. Thus, $p = 1$ is adopted in the following derivation of MTL model. In Eq. (13), $\ell_p - \ell_q (p = 1, q = 1)$ means that multi-task learning problem boils down to be L independent sub-problems. $\ell_p - \ell_q (p = 1, q > 1)$ denotes that different tasks are jointly learned. In this paper, we focus on the relationships among different tasks and only $\ell_1 - \ell_q (q > 1)$ is studied. The optimization problem of MKL and MTL (MKMTL) can be formulated as follows

$$\begin{cases} \min\limits_{\mathbf{d}} \ J(\mathbf{d}) = \sum\limits_{t=1}^{L} J_t(\mathbf{d}) \\ s.t. \ \ d_{t,m} \geq 0, \ \sum\limits_{m=1}^{M} (\sum\limits_{t=1}^{L} d_{t,m}^{q/(2-q)})^{(2-q)/q} \leq 1 \end{cases}, \tag{14}$$

with

$$\begin{cases} J_t(d) = \min\limits_{f_t} C \sum\limits_{m=1}^{M} H(f_{m,t}, y_{m,t}) + \sum\limits_{m=1}^{M} (\frac{\|f_{m,t}\|_{H_k}^2}{d_{m,t}}) \\ = \max\limits_{\alpha_t} \ \sum\limits_{i=1}^{N_t} \alpha_{t,i} - \frac{1}{2} \sum\limits_{i=1}^{N_t} \sum\limits_{j=1}^{N_t} \alpha_{t,i}\alpha_{t,j} y_{t,i} y_{t,j} \sum\limits_{m=1}^{M} d_{t,m} K_{t,m}(x_i, x_j), \\ s.t. \quad \sum\limits_{i=1}^{N_t} \alpha_{t,i} y_{t,i} = 0, \ \ 0 \leq \alpha_{t,i} \leq C, \ \ i = 1, 2 \ldots N_t \end{cases} \tag{15}$$

where $\{\alpha_t\}$ are the vectors of Lagrangian multipliers.

The problem (15) can be alternatively solved with a block-coordinate descent method [8, 20]. When keeping the matrix d, problem (15) consists of L single-task SVM sub-problems. For the task t, the objective function as following can be easily solved

$$\begin{cases} \max\limits_{\alpha_t} \ \sum\limits_{i=1}^{N_t} \alpha_{t,i} - \frac{1}{2} \sum\limits_{i=1}^{N_t} \sum\limits_{j=1}^{N_t} \alpha_{t,i}\alpha_{t,j} y_{t,i} y_{t,j} \sum\limits_{m=1}^{M} d_{t,m} K_{t,m}(x_i, x_j) \\ s.t. \quad \sum\limits_{i=1}^{N_t} \alpha_{t,i} y_{t,i} = 0, \ \ 0 \leq \alpha_{t,i} \leq C, \ \ i = 1, 2 \ldots N_t \end{cases} .. \tag{16}$$

When $\{f_{m,t}\}$ being fixed, a closed-form solution $d_{m,t}$ of this problem can be obtained [20]

$$d_{t,m} = \frac{\|f_{t,m}\|_{H_k}^{2-q} \left(\sum_v \|f_{v,m}\|_{H_k}^q \right)^{\frac{q-1}{q}}}{\left(\sum_u \left(\sum_v \|f_{v,u}\|_{H_k}^q \right)^{\frac{1}{q}} \right.}. \tag{17}$$

2.3 Radar Target Recognition with Data-Dependent MKMTL Framework

Based on MTL, a MKMTL framework is proposed in this paper. In this framework, multi-kernel learning aims to find an optimal convex combination of different kernel functions and multi-task learning utilize the shared relationships among multiple tasks to improve the classification performance of model. The proposed framework is shown in Fig. 1.

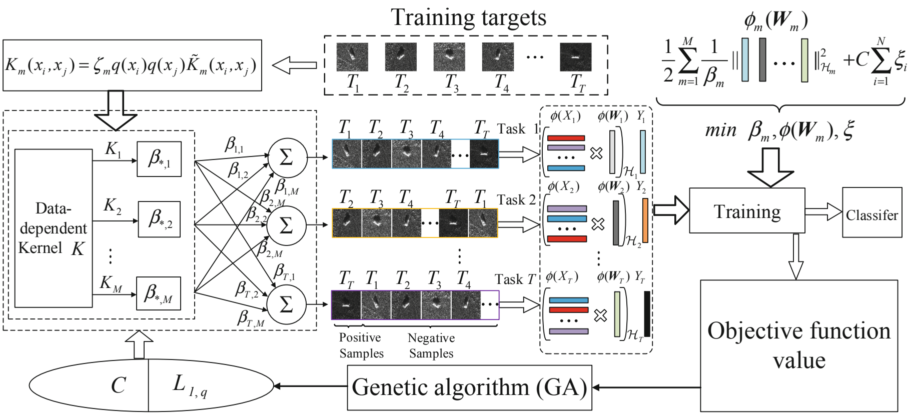

Fig. 1. The framework of proposed MKMTL.

Figure 1 shows the realization process of MKMTL model training. The whole process can be divided into several steps. Firstly, L targets are concatenated as shown in Fig. 1. In this paper, a multiple kernels learning is regarded as an independent task. Based on the concatenated target, the data-dependent kernels are calculated according to Eq. (2), where each kernel has a different bandwidth δ. In other words, the training samples are mapped into different reproducing Kernel Hilbert spaces. After that, the weighted kernel function $K_m(x_i, x_j)$ can be obtained by using Eq. (10). Then, multiple decision functions $f_t(x)$ are jointly learned based on a $\ell 1 - \ell q$ mixed norm by Eq. (14). It can be seen that all of these tasks share the same kernels sources, thus the M kernels should be able to interpret the training samples well.

In each task learning, the decision function is obtained by a multi-kernel learning. As shown in Fig. 1, the essence of multi-kernel learning is the choice of different coefficients $d_{*,i}$ according to different kernels. The multi-kernel and multi-task model training will be completed after solving problem (3). Based on the trained mode parameters, the object function value is calculated. A smaller objective function value means a better classification ability of the trained model. In the genetic algorithm, a smaller function value is mapped to a higher fitness. According to the fitness, a combination of C and q in the given range is searched. Then the new combination of C and q are assigned to multi-kernel and multi-task model and a new round of model training is initiated. This is repeated until the maximum genetic iterations is reached. With the nonlinear predictive function Eq. (12), the decision function is reached finally.

3 Experiments and Results Analysis

MSTAR public databases are consisted of SAR images of some targets. Ten targets are to be recognized in the following experiments, and Fig. 2 shows their optical and the corresponding SAR images. The original sizes of SAR images are closed to 128×128 pixels, which are cropped to 64×64 pixels in order to avoid the background clutter. The amplitude of the raw SAR image is used as an input feature. In this paper, the input feature of raw image is reduced to 200 by PCA method. In the following experiments, samples obtained at the operating condition of 17º are adopted to train the classifiers and the ones acquired at the operating condition of 15º are utilized to test the classifiers. The number of training and testing samples are shown in Table 1.

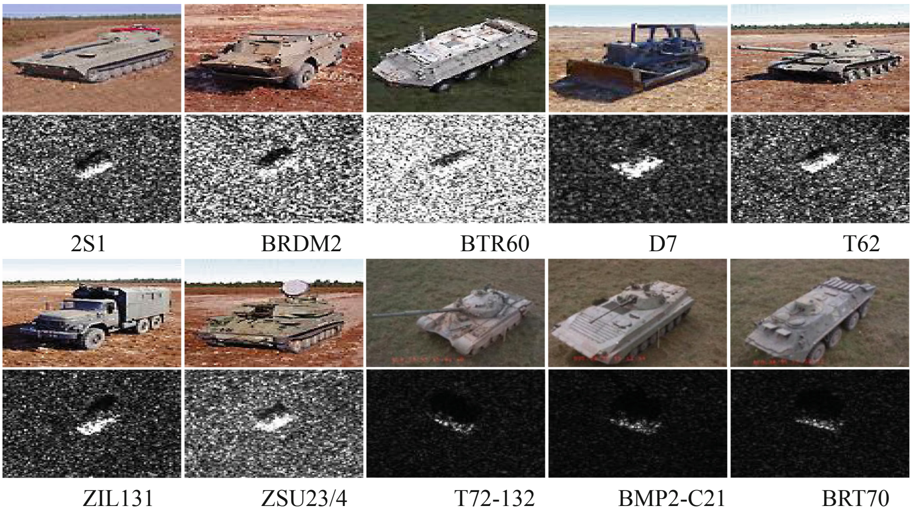

Fig. 2. The optical and SAR images of ten targets to be recognized.

Table 1. The number of training and testing samples in the ten targets.

Target	2S1	BRDM2	BTR60	D7	T62	ZIL131	ZSU23/4	BRT70	T72	BMP
Training (17°)	299	298	256	299	299	299	299	233	691	698
Testing (15°)	274	274	195	274	273	274	274	196	582	587

In the following simulation experiments, the number of kernel functions is set as 5. In CNN method, the whole net contains 4 trunk layers, including 2 convolutional layers, 1 mean pooling layer and 1 fully-connected layer. The output of the last fully-connected layer is fed to a softmax that plays a classifier role. In two convolution layers, the number of kernels is set as 20 and 120 respectively with size 13 × 13. In the pooling layer, the sampling scale is 4 × 4. In addition, the sample set is augmented by rotating the image clockwise by 60°, 120°, 180° and 240° respectively. The number of iterations of CNN is designed as 100. In this experiment, three targets recognition experiment, the vehicles ‘2S1’, ‘BRDM2’ and ‘ZSU23/4’ are selected as the targets and the recognition results are shown in Fig. 3 (Table 2).

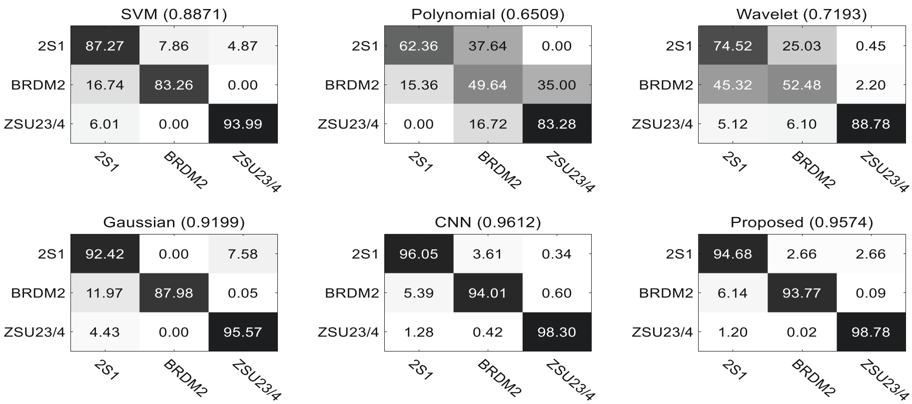

Fig. 3. Confusion matrices of the compared methods.

Table 2. Calculations of different methods.

Methods	Number of multiplications	Number of additions
SVM	$N \cdot n$	$2N \cdot n - 1$
Polynomial	$M \cdot N \cdot (n + l)$	$2M \cdot N \cdot n - 1$
Wavelet	$M \cdot N \cdot [n(n+2) - 1]$	$M \cdot N \cdot [n(3n-1)+1] - 1$
Gaussian	$M \cdot N \cdot n$	$2M \cdot N \cdot n - 1$
Proposed	$M \cdot N \cdot [n(k+1)+1]$	$M \cdot N \cdot [2n(k+1) - 1] - 1$
CNN	$C_i^2 \cdot M_{i-1} \cdot M_i \cdot [(I_i^h - C_i + 1)/S_i] \cdot [(I_i^w - C_i + 1)/S_i]$	$(C_i^2 \cdot M_{i-1} - 1) \cdot M_i \cdot [(I_i^h - C_i + 1)/S_i] \cdot [(I_i^w - C_i + 1)/S_i]$

Figure 3 indicates that the average recognition accuracy for the proposed method is 0.9574, compared to 0.8817 for SVM, 0.6509 for Polynomial kernel, 0.7193 for Wavelet kernel, and 0.9199 for Gaussian kernel. It is 7.57%, 30.65%, 23.81%, 23.72% and 3.75% better than the competitors, KNN, SVM, Polynomial kernel, Wavelet kernel, and Gaussian kernel, respectively. These numerical results prove that different kinds of data can be effectively mapped into different RKHSs through the weighted data-dependent kernel, which is generated by the separable measure and Kernel Alignment measure criteria. When compared with CNN deep learning method, the performance of proposed method is slightly worse than CNN. But the complexity of CNN method is much higher than MKMTL method. It can be concluded that the recognition performance of the proposed method is close to deep learning CNN while the calculation amount is lower than the deep learning. To further test the performance of the proposed MKMTL method, the classification experiment of ten targets is implemented and the results are shown in Fig. 4.

Figure 4 shows that CNN and proposed method achieve the ideal recognition rate. The performance of the proposed MKMTL method is 4.69%, 15.42%, 9.71% and

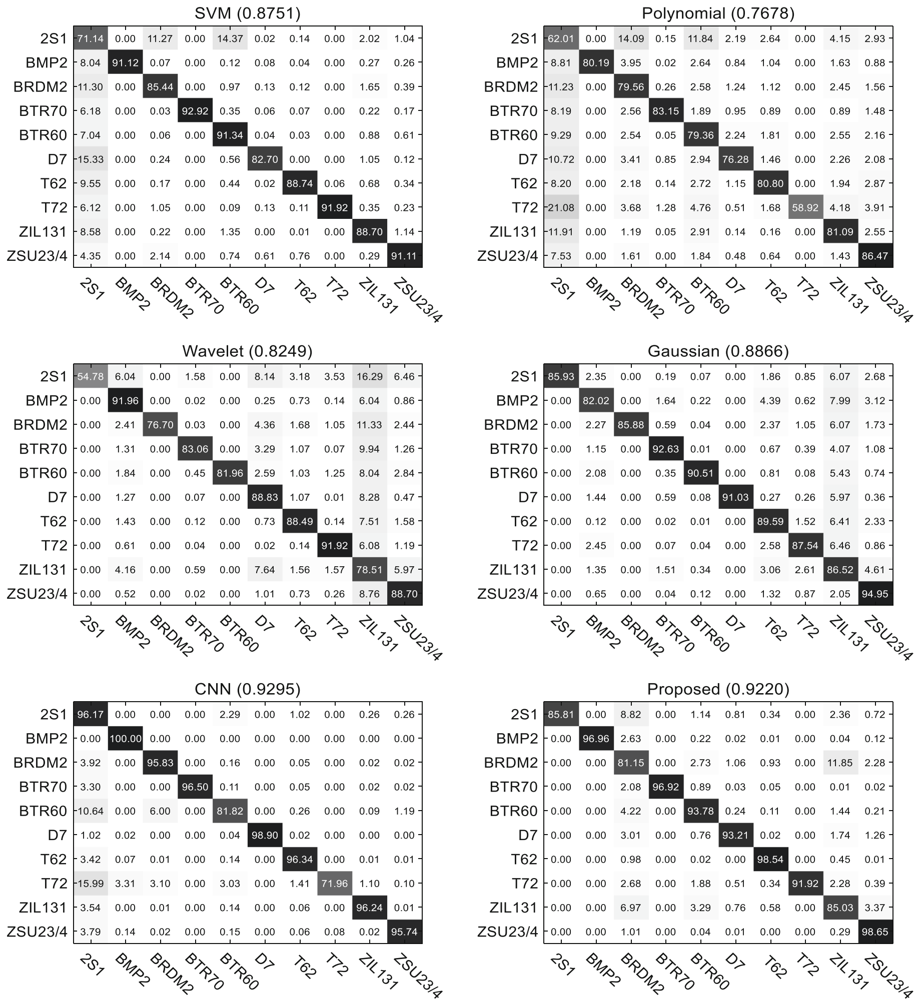

Fig. 4. Confusion matrices of the compared methods.

3.54% better than the competitors, SVM, Polynomial kernel, Wavelet kernel, and Gaussian kernel, respectively. The experimental results show that MKMTL method can interpret the training samples well and fully reveal the relationships among ten tasks. Polynomial kernel and wavelet kernel based methods have a better performance in the case of ten targets recognition. It indicates that more and more information among multiple tasks can be shared with the number of tasks increasing, which improves the performance of model.

4 Conclusion

In this paper, a data-dependent MKMTL framework is proposed to realize the radar target recognition. In order to improve the interpretability of input data, a weighted data-dependent kernel is proposed, which can adapt to different data sets and improve the classification performances. To accurately describe different data sets, the proposed kernel is extended to multiple kernels. Furthermore, multi-task learning framework is proposed to fully reveal and utilize the hidden information among different tasks. To increase the flexibility of multi-task model, a larger class of mixed norm penalty is used. Experiments testify the superiority and practicability of the proposed MKMTL method.

References

1. Liu, A.N., et al.: HEp-2 cells classification via clustered multi-task learning. Neurocomputing **195**(C), 195–201 (2016)
2. Kang, M., et al.: Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder. Sensors **17**(1), 192 (2017)
3. Wang, S.N., et al.: SAR image target recognition via complementary spatial pyramid coding. Neurocomputing **196**(C), 125–132 (2016)
4. Odysseas, K.S., Nabil, A.: Fusing deep learning and sparse coding for SAR ATR. IEEE Trans. Aerosp. Electron. Syst. **55**(2), 785–797 (2019)
5. Tian, Z., et al.: Classification via weighted kernel CNN: application to SAR target recognition. Int. J. Remote Sens. **39**(3), 1–20 (2018)
6. Su, H.J., et al.: Kernel collaborative representation with local correlation features for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **57**(2), 1230–1241 (2019)
7. Wei, C.C.: Wavelet kernel support vector machines forecasting techniques: case study on water-level predictions during typhoons. Expert Syst. Appl. **39**(5), 5189–5199 (2012)
8. Rakotomamonjy, A., et al.: SimpleMKL. J. Mach. Learn. Res. **9**(3), 2491–2521 (2008)
9. Micchelli, C.A., Pontil, M.: Learning the kernel function via regularization. J. Mach. Learn. Res. **6**(6), 1099–1125 (2005)
10. Xu, Y.Y., et al.: An efficient renovation on kernel fisher discriminant analysis and face recognition experiments. Pattern Recogn. **37**(10), 2091–2094 (2004)
11. Wang, Q.W., Gu, Y.F., Tuia, D.: Discriminative multiple kernel learning for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **54**(7), 3912–3927 (2016)

12. Gu, Y.F., et al.: Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **8** (6), 2739–2751 (2015)
13. Gu, Y.F., Wang, Q.W., Xie, B.Q.: Multiple kernel sparse representation for airborne LiDAR data classification. IEEE Trans. Geosci. Remote Sens. **PP**(99), 1–21 (2016)
14. Karine, A., et al.: Radar target recognition using salient keypoint descriptors and multitask sparse representation. Remote Sens. **10**(6), 843 (2018)
15. Zhang, X.Z., et al.: Two-stage multi-task representation learning for synthetic aperture radar (SAR) target Images classification. Sensors. **17**(11), 2506 (2017)
16. Ozawa, S., Roy, A., Roussinov, D.: A multitask learning model for online pattern recognition. IEEE Trans. Neural Networks **20**(3), 430–445 (2009)
17. Du, L., et al.: Bayesian spatiotemporal multitask learning for radar HRRP target recognition. IEEE Trans. Signal Process. **59**(7), 3182–3196 (2011)
18. Zheng, H., et al.: Multi-task model for simultaneous face identification and facial expression recognition. Neurocomputing. **171**(C), 515–523 (2016)
19. Dong, G.G., et al.: SAR target recognition via joint sparse representation of monogenic signal. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **8**(7), 3316–3328 (2015)
20. Rakotomamonjy, A., et al.: Penalty for sparse linear and sparse multiple kernel multitask learning. IEEE Trans. Neural Networks **22**(8), 1307–1320 (2011)
21. Lounici, K., et al.: Taking advantage of sparsity in multi-task Learning. In: COLT 2009 (2009)
22. Xiong, H., Swamy, M.N., Ahmad, M.O.: Optimizing the kernel in the empirical feature space. IEEE Trans. Neural Networks **16**(2), 460–474 (2005)
23. Cristianini, N., et al.: On kernel-target alignment. Adv. Neural. Inf. Process. Syst. **179**(5), 367–373 (2002)