Article

Clustered Multi-Task Learning for Automatic Radar Target Recognition

Cong Li, Weimin Bao, Luping Xu * and Hua Zhang

School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China; lcongxd@126.com (C.L.);
baoweimin@cashq.ac.cn (W.B.); zhanghua@mail.xidian.edu.cn (H.Z.)
* Correspondence: lpxu@mail.xian.edu.cn; Tel.: +86-029-8189-1862

Abstract: Model training is a key technique for radar target recognition. Traditional model training algorithms in the framework of single task leaning ignore the relationships among multiple tasks, which degrades the recognition performance. In this paper, we propose a clustered multi-task learning, which can reveal and share the multi-task relationships for radar target recognition. To further make full use of these relationships, the latent multi-task relationships in the projection space are taken into consideration. Specifically, a constraint term in the projection space is proposed, the main idea of which is that multiple tasks within a close cluster should be close to each other in the projection space. In the proposed method, the cluster structures and multi-task relationships can be autonomously learned and utilized in both of the original and projected space. In view of the nonlinear characteristics of radar targets, the proposed method is extended to a non-linear kernel version and the corresponding non-linear multi-task solving method is proposed. Comprehensive experimental studies on simulated high-resolution range profile dataset and MSTAR SAR public database verify the superiority of the proposed method to some related algorithms.

Keywords: clustered multi-task learning; high-resolution range profile (HRRP); synthetic aperture radar (SAR); radar automatic target recognition (RATR)

---

1. Introduction

Automatic target recognition (ATR) systems are used to identify one or a group of target objects in a scene. These ATR systems are to detect and classify targets using various images and signal processing techniques. Due to the capability of producing all-weather, 24-h a day and robustness towards the environmental condition of radar sensors, researchers have drawn much attention on the automatic target recognition based on radar images. Usually, radar images can be divided into one-dimensional high-resolution range profile (HRRP) images and two-dimensional images, like synthetic aperture radar (SAR) images. In recent years, radar images have been intensively studied for ATR in civilian and military fields [1–7]. Nevertheless, HRRPs and SAR images are sensitive to the variation in the pose and the speckle noise. How to recognize the specified radar targets still requires further study and exploration.

Radar target recognition generally consists of three main separate stages: detection, discrimination, and classification. The first stage aims to approximately determine the location of targets by using the amplitude information of radar signals. The second stage excludes the interference of clutter in background. The last stage is to predict the category of targets using classifiers. In this paper, the classifier design is emphasized. Lots of classical classifiers have been implemented for radar target classification, including k-nearest neighbors (KNN), support vector machine (SVM) [8], AdaBoost [9] and so on. Zhou incorporated the reconstructive power and discriminative power of dictionary atoms for radar target recognition [2]. A complementary spatial pyramid coding (CSPC) approach for radar

target recognition was realized by Wang et al. [4]. Song et al. came up with a supervised discriminative dictionary learning (SDDL) method by learning a discriminative dictionary for SAR ATR [6].

Although much work has been done on radar target recognition, two critical problems still exist. To our knowledge, most of the previous methods have been implemented in the framework of single-task learning. Although many sophisticated classifiers have been designed, they cannot make full use of the latent relatedness among multiple radar categories. Therefore, it is challenging to classify the targets with similar patterns by single-task learning methods. On the other hand, the acquired samples are usually limited or imbalanced, since large-scale and complete dataset collection is expensive. Single-task learning can't realize information (common structures or parameters) sharing, which makes it difficult to train a model. On the contrary, multi-task learning can make use of multi-task relationships and thus help to improve the recognition accuracy. Dong et al. considered three components of monogenic signals as different learning tasks and proposed a joint sparse representation model to exploit the intercorrelation among multiple tasks for SAR ATR [10]. The contrast experiments have proven the superiority of multi-task learning.

A few studies have focused on the multi-task learning for RATR except for [10], even though multi-task learning is superior to single task leaning. Inspired by the recent successful application of multi-task learning in the field of ATR, such as facial expression recognition [11], HEp-2 cell classification [12], and Alzheimer's disease (AD) diagnosis classification [13], we propose a new classification method based on clustered multi-task learning theory. Two issues should be taken into consideration when the clustered multi-task learning method is used for radar target recognition. Due to the target-aspect and pose sensitivity of radar images, the scattering characteristics among different targets in the same target-aspect may be extremely similar, which makes it difficult for humans to cluster radar targets simply using visual appearance. The other problem is that the geometric structure information hidden in the radar images, such as target size and scatterer distribution is complicated and nonlinear, which results in a difficult extraction of cluster information. The multi-task relationship learning (MTRL) method proposed in [14] can autonomously learn the positive and negative task correlations, and also can be easily extended to nonlinear domains, which can address these two issues well. To further improve the classification ability of MTRL, a projection regularization term is added into the objective function to fully explore the cluster structure and multi-task relationships in the original and projected space of model parameter. The projection term assumes that the multiple tasks within a close cluster should be close to each other in the projection space. As an extension of the MTRL method, the proposed method can autonomously learn multi-task relationships, cluster the information of different tasks and be easily extended to nonlinear domains. The main difference between the proposed method and the approach in [10] is that the former can automatically learn the multi-task relationships and clusters hidden in the kernel function space. In terms of object function optimization, SMO method [15] is frequently adopted in the non-linear versions of single-task learning. However, when SMO is used for solving the non-linear extension of multi-task learning, multiple variables need to be processed simultaneously, which greatly increases the heuristic selection time and sometimes can't guarantee convergence. To overcome this problem, the objective function is firstly transformed into a dual form, and then the widely applied APG method [16] is adopted to solve it. The proposed solving method can guarantee the convergence and be implemented in parallel computing. In addition, extensive studies on the simulated and real databases are conducted to assess and analyze the performance of the proposed method. The proposed method includes the following advantages:

(1) The theory of clustered multi-task learning is applied to radar target recognition.
(2) The potentially useful multi-task relationships in the projection space are taken into consideration, which helps to discriminate the radar targets with similar patterns.
(3) The proposed method can autonomously learn the multi-task relationships, cluster information and be easily extended to nonlinear domains.

(4)　APG method is used for solving the non-linear extension of multi-task learning, which guarantees the convergence and can be implemented in parallel computing.

The rest of the paper is organized as follows: in Section 2, the clustered multi-task learning for radar target recognition is proposed. The experimental results and analysis are provided in Section 3, and the paper is finalized with conclusions in Section 4.

2. Clustered Multi-Task Learning

2.1. Preliminaries

For radar target classification, the model learning for m target categories can be considered as m tasks. During the training phase, a set of training samples $\left\{(x_j^i, y_j^i)\right\}_{j=1}^{n_i}$ are given for model learning, where $x_j^i \in R^d$ is the d-dimension training data, $y_j^i \in \{0,1\}$ is the label and n_i is the number of samples in the ith task. In this paper, the goal is to learn a nonlinear predictive function $f_i(x_j^i) = \phi(x_j^i)^T w_i + b_i$, where $\phi(x_j^i)$ is the nonlinear mapping of sample x_j^i, w_i is the model parameter and b_i is the offset of ith task. Let $\mathbf{W} = [w_1 \dots w_m]$, then the objective function can be formulated as:

$$\min_{\mathbf{W}} f(\mathbf{W}) + \Omega(\mathbf{W}) \tag{1}$$

where $f(\mathbf{W})$ is the empirical loss function. In this paper, it is defined as:

$$f(\mathbf{W}) = \sum_{i=1}^{m} \frac{1}{n_i} \sum_{j=1}^{n_i} \left(y_j^i - \phi(x_j^i)^T w_i - b_i\right)^2 \tag{2}$$

where n_i is to alleviate the data imbalance among different tasks. $\Omega(\mathbf{W})$ is clustering-based regularization to constrain the shared information among different tasks, and has been the focus of many researchers.

For example, the authors assume that the parameter vector of each task is similar to the average parameter vector, and $\Omega(\mathbf{W})$ is formulated as [17]:

$$\Omega(\mathbf{W}) = \sum_{i=1}^{m} \sum_{k=1}^{m} \frac{1}{2m} \|w_i - w_k\|_2^2 = tr(\mathbf{WLW}^T) \tag{3}$$

where L is the Laplacian matrix defined on the graph with edge weights equaling to $1/2m$. In [18], the authors assume that all tasks can be grouped into $r < m$ clusters, and $\Omega(\mathbf{W})$ is defined as:

$$\Omega(\mathbf{W}) = \sum_{c=1}^{r} \sum_{v \in I_c} \|w_v - \overline{w}_c\|_2^2 = tr(\mathbf{W}^T \mathbf{W}) - tr(\mathbf{F}^T \mathbf{W}^T \mathbf{W} \mathbf{F}) \tag{4}$$

where I_c is the index set of the cth cluster, $\overline{w}_j$ denotes the mean of the cth cluster, and matrix $\mathbf{F} \in \mathbf{R}^{m \times r}$ is an orthogonal cluster indicator matrix with $F_{i,c} = 1/\sqrt{n_c}$ if $i \in I_c$ and $F_{i,c} = 0$ otherwise. These methods assume that the cluster structures or the multi-task relationships are known. Nevertheless, sometimes these model assumptions may be incorrect or even worse. Thus, learning the task relationships from data automatically is a more favorable choice. In [14], a multi-task relationship learning (MTRL) method is proposed, which can autonomously learn the positive and negative task correlation. The $\Omega(\mathbf{W})$ is given as:

$$\Omega(\mathbf{W}) = tr(\mathbf{WS}^{-1}\mathbf{W}^T) \quad s.t. \quad \mathbf{S} \succeq 0, \, tr(\mathbf{S}) = 1 \tag{5}$$

where $\mathbf{S}$ is defined as a task covariance matrix and $\mathbf{W}$ is the model parameter.

2.2. Proposed Clustered Multi-Task Learning

In MTRL method, the multi-task relationships among different mode parameters are fully utilized. To further improve the classification performance of MTRL, we assume that the tasks with a close relationship should be close to each other in the projection space $\mathbf{X}^T\mathbf{W}$. That is to say, the task covariance matrix $\mathbf{S}$ should reflect the multi-task relationships in the original and projected space of mode parameters. The proposed $\Omega(\mathbf{W})$ can be formulated as:

$$\Omega(\mathbf{W}) = \frac{\lambda_1}{2}\|\mathbf{W}\|_2^2 + \frac{\lambda_2}{2}tr(\mathbf{W}\mathbf{S}^{-1}\mathbf{W}^T) + \frac{\lambda_3}{2}tr(\mathbf{P}\mathbf{S}^{-1}\mathbf{P}^T)$$
$$s.t.\quad \mathbf{S}\succeq 0,\ tr(\mathbf{S}) = 1,\ p_j^i = (x_j^i)^T w_i \tag{6}$$

where $\mathbf{P} = [P_1,\ldots,P_m]$, $P_i = [p_1^i,\ldots,p_{n_i}^i]$, λ_i is the regularization parameter, and $\mathbf{S}$ denotes the task covariance matrix. The target features hidden in the radar images are usually nonlinear. Thus a nonlinear kernel version of the proposed method is obtained:

$$\Omega(\mathbf{W}) = \frac{\lambda_1}{2}\|\mathbf{W}\|_2^2 + \frac{\lambda_2}{2}tr(\mathbf{W}\mathbf{S}^{-1}\mathbf{W}^T) + \frac{\lambda_3}{2}tr(\mathbf{P}\mathbf{S}^{-1}\mathbf{P}^T) \tag{7}$$

where $\mathbf{P} = [P_1,\ldots,P_m]$, $P_i = \phi(X_i)^T w_i$, and $\phi(X_i) = [\phi(x_1^i),\ldots,\phi(x_{n_i}^i)]$. The first term penalizes the complexity of $\mathbf{W}$. The second term restricts the distance between w_i and w_k in the model parameter space, and the third term controls the distance between $\phi(X_i)$ and $\phi(X_k)$ in the projected space. The latter two regularization terms imply that the distance between a pair of task T_i and T_k should be as small as possible in the original and projected space if they belongs to the same cluster. To sum up, the objective function can be denoted as:

$$\min_{W,b}\sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}(\xi_j^i)^2 + \frac{\lambda_1}{2}\|\mathbf{W}\|_2^2 + \frac{\lambda_2}{2}tr(\mathbf{W}\mathbf{S}^{-1}\mathbf{W}^T) + \frac{\lambda_3}{2}tr(\mathbf{P}\mathbf{S}^{-1}\mathbf{P}^T)$$
$$s.t.\quad \xi_j^i = y_j^i - \phi(x_j^i)^T w_i - b_i,\quad p_j^i = \phi(x_j^i)^T w_i \tag{8}$$

2.3. Proposed Optimization Method

The objective function of problem (8) is convex on all variables. But it is not easy to optimize the objective function with respect to all the variables simultaneously. Here an alternating method is adopted to solve the problem. Firstly, $\mathbf{W}$ and $\mathbf{b}$ are updated with fixed $\mathbf{S}$. Then $\mathbf{S}$ is updated with fixed $\mathbf{W}$ and $\mathbf{b}$.

Specifically, when $\mathbf{S}$ is fixed, the optimization problem for updating $\mathbf{W}$ and $\mathbf{b}$ can be stated as:

$$\min_{W,b}\sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}(\xi_j^i)^2 + \frac{\lambda_1}{2}\|\mathbf{W}\|_2^2 + \frac{\lambda_2}{2}tr(\mathbf{W}\mathbf{S}^{-1}\mathbf{W}^T) + \frac{\lambda_3}{2}tr(\mathbf{P}\mathbf{S}^{-1}\mathbf{P}^T)$$
$$s.t.\quad \xi_j^i = y_j^i - \phi(x_j^i)^T w_i - b_i,\quad p_j^i = \phi(x_j^i)^T w_i \tag{9}$$

To facilitate a kernel extension for proposed method, the optimization problem is transformed into a dual form:

$$L = \sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}(\xi_j^i)^2 + \frac{\lambda_1}{2}\|\mathbf{W}\|_2^2 + \frac{\lambda_2}{2}tr(\mathbf{W}\mathbf{S}^{-1}\mathbf{W}^T) + \frac{\lambda_3}{2}tr(\mathbf{P}\mathbf{S}^{-1}\mathbf{P}^T)$$
$$+ \sum_{i=1}^{m}\sum_{j=1}^{n_i}\alpha_j^i(y_j^i - \phi(x_j^i)^T w_i - b_i - \xi_j^i)^2 + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\beta_j^i(p_j^i - \phi(x_j^i)^T w_i)^2 \tag{10}$$

where α^i_j and β^i_j are the Lagrange multipliers associated with the jth training sample of the ith task. Setting the derivative of L with respect to W, b_i, ξ^i_j and P equal to zero, we obtain:

$$
\begin{aligned}
\frac{\partial L}{\partial \mathbf{W}} &= \lambda_1 \cdot \mathbf{W} + \lambda_2 \cdot \mathbf{W} \mathbf{S}^{-1} - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha^i_j \phi(x^i_j) e_i^T - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \beta^i_j \phi(x^i_j) e_i^T = 0 \\
\frac{\partial L}{\partial b_i} &= -\sum_{j=1}^{n_i} \alpha^i_j = 0 \\
\frac{\partial L}{\partial \xi^i_j} &= \frac{2}{n_i}\xi^i_j - \alpha^i_j = 0 \\
\frac{\partial L}{\partial P} &= \lambda_1 \cdot \mathbf{P} \mathbf{S}^{-1} + \mathbf{B} = 0
\end{aligned}
\tag{11}
$$

where $\mathbf{e}_i$ is the ith column vector of $I_{m \times m}$ and β^i_j is the (j,i)th element of $\mathbf{B}$. Plugging Equation (11) into Equation (10), the following form is obtained

$$
E = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha^i_j y^i_j - \frac{n_i}{4}\sum_{i=1}^{m}\sum_{j=1}^{n_i} (\alpha^i_j)^2 - \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\beta}^T \mathbf{K} \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\beta}^T \mathbf{K} \boldsymbol{\beta} - \frac{1}{2\lambda_3}\boldsymbol{\beta}^T \widetilde{\mathbf{S}} \boldsymbol{\beta} \quad (12)
$$

where $\boldsymbol{\alpha} = (\alpha^1_1, \ldots, \alpha^m_n)^T$ and $\boldsymbol{\beta} = (\beta^1_1, \ldots, \beta^m_n)^T$. $\mathbf{K} \in \mathbf{R}^{\sum_{i=1}^{m} n_i \times \sum_{i=1}^{m} n_i}$ is the multi-task kernel matrix defined on all the training samples. For any two training samples $(x^{i_1}_{j_1}, x^{i_2}_{j_2})$, the corresponding multi-task kernel is referred to as $K(x^{i_1}_{j_1}, x^{i_2}_{j_2}) = \mathbf{e}_{i_1}^T (\lambda_1 I_{m \times m} + \lambda_2 L)^{-1} \mathbf{e}_{i_2} \phi(x^{i_1}_{j_1})^T \phi(x^{i_2}_{j_2})$. $\widetilde{\mathbf{S}} \in \mathbf{R}^{\sum_{i=1}^{m} n_i \times \sum_{i=1}^{m} n_i}$ contains $m \times m$ sub-matrices and the (i,j)th sub-matrix is a diagonal matrix with diagonal element $\mathbf{S}(i,j)$. Since problem (12) is an unconstrained optimization problem with respect to $\boldsymbol{\beta}$, the derivative of E with respect to $\boldsymbol{\beta}$ can be given as:

$$
\frac{\partial E}{\partial \boldsymbol{\beta}} = -\mathbf{K}\boldsymbol{\alpha} - \mathbf{K}\boldsymbol{\beta} - \frac{\widetilde{\mathbf{L}}}{\lambda_3}\boldsymbol{\beta}
\tag{13}
$$

Setting the derivative equal to zero, we obtain:

$$
\boldsymbol{\beta} = \mathbf{Q}^{-1}\mathbf{K}\boldsymbol{\alpha}
\tag{14}
$$

where $\mathbf{Q} = -(\mathbf{K} + \frac{\widetilde{\mathbf{S}}}{\lambda_3})$. Substituting Equation (14) into Equation (12), we can obtain:

$$
\min_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T \widetilde{\mathbf{K}} \boldsymbol{\alpha} - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha^i_j y^i_j
\tag{15}
$$

where $\widetilde{\mathbf{K}} = \mathbf{K} + 2\mathbf{K}\mathbf{Q}^{-1}\mathbf{K} + \mathbf{K}\mathbf{Q}^{-1}\mathbf{K}\mathbf{Q}^{-1}\mathbf{K} + 1/\lambda_3 \mathbf{K}\mathbf{Q}^{-1}\widetilde{\mathbf{L}}\mathbf{Q}^{-1}\mathbf{K} + 1/2\Lambda$, and $\Lambda \in \mathbf{R}^{\sum_{i=1}^{m} n_i \times \sum_{i=1}^{m} n_i}$ is a diagonal matrix with diagonal element n_i if the corresponding data point is from the ith task.

So far, the problem (9) is converted to a familiar form. In most literatures [14,19], Equation (15) is solved by an SMO algorithm, being similar to the least-squares SVM [15]. However, multiple variables need to be heuristically selected, when SMO method is used for solving the non-linear extension of multi-task learning. In this paper, the SMO approach can't guarantee convergence due to the mutual interference when multiple variables are heuristically selected. In the proposed solving method, the problem (15) is efficiently solved by the widely applied APG method [16]. Specifically, the objective function is divided into the smooth part $f(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^T \widetilde{\mathbf{K}} \boldsymbol{\alpha}$ and non-smooth part $g(\boldsymbol{\alpha}) = -\sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha^i_j y^i_j$.

The gradient of $f(\boldsymbol{\alpha})$ is Lipschitz continuous with the Lipschitz constant ℓ satisfying with $\ell \geq \|\widetilde{\mathbf{K}}\|_2$. Given ℓ, a surrogate function $T_\ell(\boldsymbol{\alpha}, \boldsymbol{\alpha}_k)$ is defined as:

$$
T_\ell(\boldsymbol{\alpha}, \boldsymbol{\alpha}_k) = f(\boldsymbol{\alpha}_k) + <\boldsymbol{\alpha} - \boldsymbol{\alpha}_k, \nabla f(\boldsymbol{\alpha}_k)> + \frac{\ell}{2}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\|_2^2 + g(\boldsymbol{\alpha})
\tag{16}
$$

where $\nabla f(\alpha_k)$ denotes the derivative of $f(\alpha)$ with respect to α at $\alpha = \alpha_k$. After we omitted the constant terms, Equation (16) can be redefined as:

$$M_\ell(\alpha_k) = \min_\alpha \frac{\ell}{2}\|\alpha - \alpha_k\|_2^2 + \alpha^T \widetilde{K}\alpha_k - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha_j^i y_j^i \quad s.t. \sum_{j=1}^{n_i} \alpha_j^i = 0 \, \forall i \tag{17}$$

which can be decoupled into m subproblems with the ith one formulated as:

$$M_\ell((\alpha_k)_i) = \min_\alpha \frac{\ell}{2}\|\alpha_i - (\alpha_k)_i\|_2^2 + \alpha_i^T z_i - \sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha_j^i y_j^i \quad s.t. \sum_{j=1}^{n_i} \alpha_j^i = 0 \tag{18}$$

where $z = \widetilde{K}\alpha_k$ and z_i is a subvector of z corresponding to the $i$thtask. It is not difficult to see that Equation (17) can be easily implemented in parallel computing. Based on the Lagrange multiplier method, we can obtain the analytical solution of problem (17) as follows:

$$\alpha = \frac{1}{\ell}(Y - D) + V \tag{19}$$

where $Y = (y_1^1, \dots, y_{n_m}^m)^T$, $V = (v_1^1, \dots, v_{n_m}^m)^T$, $v^i = (v_1^i, \dots, v_{n_i}^i)^T = (\alpha_k)_i - z_i/\ell$, $D = (d_1^1, \dots, d_{n_m}^m)^T$, and $d_j^i = \frac{1}{n_i}\sum_{j=1}^{n_i} \ell \cdot v_j^i + y_j^i$.

When W and b are obtained, the subproblem for minimizing Equation (10) over S can be stated as:

$$\min_S tr(\lambda_2 tr(WS^{-1}W^T) + \lambda_3 tr(PS^{-1}P^T)) \tag{20}$$
$$s.t.\, S \succeq 0, \, tr(S) = 1$$

Similar to [14], the analytical solution S is given as:

$$S = \frac{(\lambda_2 W^T W + \lambda_3 P^T P)^{1/2}}{tr((\lambda_2 W^T W + \lambda_3 P^T P)^{1/2})} \tag{21}$$

Then the nonlinear predictive function $f_i(x_j^i) = \phi(x_j^i)^T w_i + b_i$ can be obtained:

$$f_i(x_j^i) = \sum_{i_0=1}^{m}\sum_{j_0=1}^{n_i} (\alpha_{j_0}^{i_0} + \beta_{j_0}^{i_0})K(x_{j_0}^{i_0}, x_j^i) + \frac{1}{n_i}\sum_{j=1}^{n_i} \left(\frac{n_i \alpha_j^i}{2} + T_j^i\right) \tag{22}$$

where $T_j^i = \alpha^T \widetilde{k}_j^i - y_j^i$ and $\widetilde{k}_j^i$ is a column of $\widetilde{K}$ corresponding to x_j^i. The proposed method can be pictorially shown in Figure 1, and the corresponding recognition procedure is given in Algorithm 1.

3. Experimental Results and Analysis

In order to verify the effectiveness and robustness of the proposed method, simulated HRRP datasets and MSTAR SAR public databases are used for tests. In the following studies, a one-against-all framework is adopted, where each one-against-all classification problem is considered as a task. To quantitatively assess the performance, several state-of-the-art algorithms, including KNN, SVM and some other multi-task learning methods, which are summarized in Table 1, are used as the reference.
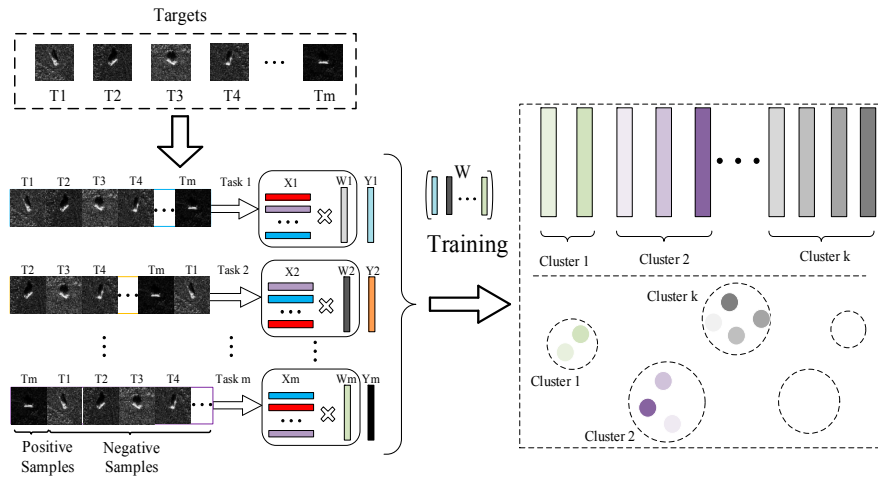
Figure 1. Illustration of proposed method. Firstly, m tasks are formed by concatenating the m targets. In each of the task, the first target is considered as positive sample, and the others negative ones. Then, multi-task learning is performed under the constraint of objective function Equation (8). After training, the multi-task relationships and clustered structures are obtained by Equation (21). In the figure, tasks with similar colors are similar with each other. Finally, the decision is reached by the obtained nonlinear predictive function Equation (22).

Algorithm 1. Pseudo Code for Solving Problem (8)

1: Input $\mathbf{X} \in \mathbf{R}^{d \times \sum_{i=1}^{m} n_i}$, $\mathbf{Y} \in \mathrm{R}^{\sum_{i=1}^{m} n_i \times 1}$, $\lambda_1, \lambda_2, \lambda_3$;
2: Initialize $\mathbf{S} = \mathbf{I}^{m \times m}$, $\mathbf{W} = \mathbf{0}^{d \times m}$ and $\mathbf{b} = \mathbf{0}^{1 \times m}$;
3: while not converged
4:     Update $\mathbf{W}$ and $\mathbf{b}$
5:         Reformulate the optimize problem (9) into a dual form (12)
6:         Update $\boldsymbol{\beta}$ by Equation (14)
7:         Solving problem (15) by using the APG method
8:     Update $\mathbf{S}$ by using Equation (21)
9: end while
10: Output $\mathbf{W}$, $\mathbf{b}$ and $\mathbf{S}$.

Table 1. The reference methods to be studied.

| Methods	Description
KNN	K-nearest neighbor classifier.
SVM [8]	Support vector machine learning.
Trace-norm Regularized multi-task learning (Trace) [20]	Trace method assumes that all models share a common low dimensional subspace.
Regularized multi-task learning (RMTL) [17]	RMTL method assumes that all tasks are similar, and the parameter vector of each task is similar to the average parameter vector.
Clustered Multi-Task Learning (CMTL) [18]	CMTL assumes that multiple tasks follow a clustered structure and that such a clustered structure is prior. In the experiments, we perform multiple single task learning to get the trained mode parameters, and based on which to obtain the clustered structure.
Multi-task relationship learning (MTRL) [14]	MTRL can autonomously learn the positive and negative task correlation.

3.1. Investigations Based on a Simulated Database

In the simulation experiments, three categories of tank models are considered as the radar targets. HRRP samples are obtained by performing an IFFT of RCS samples, which are generated by FEKO software, whose electromagnetic simulation parameters are listed in Table 2. The three targets and their corresponding HRRPs are shown in Figure 2. In the simulation, the profiles generated at the depression angle of 15° are randomly divided into two equal parts, one for training and the other for testing. For each target, a 50-dimensional feature is achieved by principal component analysis (PCA). In this section, the influences of model parameters on the recognition performance are discussed, then two comparative experiments are conducted. In each of the numerical simulation, twenty times Monte Carlo simulation experiments are performed to achieve the average recognition results.

Table 2. Electromagnetic simulation parameters.

Waveform	Center Frequency	Band-Width	Number of Frequency Samples	Meshing Size	Depression Angles	Azimuth Angles
Chirp Signal	1.5 GHz	1 GHz	1000	$\lambda/6$	15°	0–180° with 1° steps
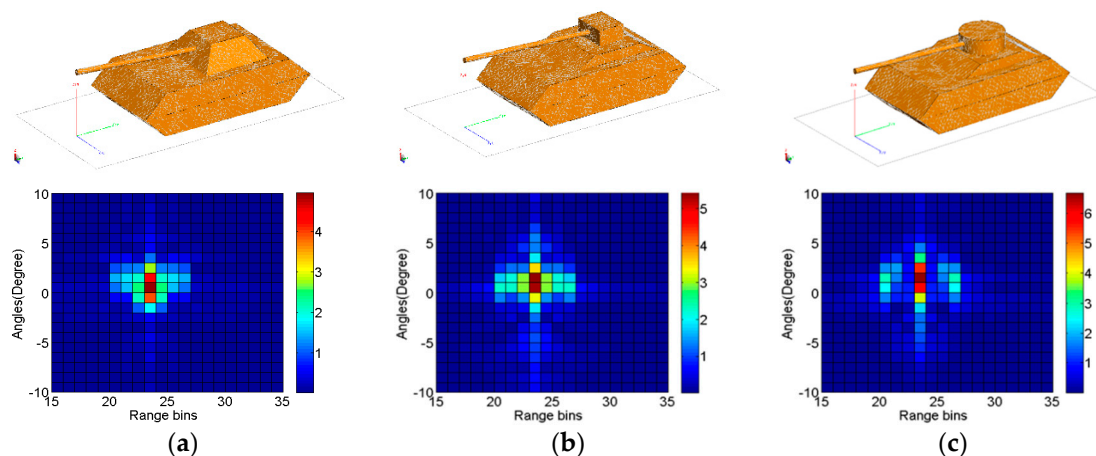
**Figure 2.** Geometrical models (top) and HRRPs (bottom) of three tanks. (**a**) Tank 1; (**b**) Tank 2; (**c**) Tank 3.

3.1.1. Influence of Model Parameters

In the prosed model, there are three parameters λ_1, λ_2 and λ_3. Here we set $\lambda_1 = 1$, and only consider the influence of parameters $\lambda_2 \in \left\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\right\}$ and $\lambda_3 \in \left\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\right\}$.

Two metrics of ACC (accuracy) and AUC (area under curve) are adopted to validate the proposed method. The average results are shown in Figure 3. Figure 3a shows that the best recognition rate is reached when λ_2 and λ_3 are equal to 10^{-1} and 10^{-3}, respectively. The optimal value of λ_3 is less than that of λ_2, when our method achieves the best performance. This indicates that the distances among three tasks in the model parameter space are larger than that in the projected space. The penalty of the distance in model parameter space is severer than that in projected space if $\lambda_2 > \lambda_3$, which promotes the generalization ability. Figure 3b shows that the value of AUC approximately equals to one in most of the combinations of λ_2 and λ_3, which means that our model has a strong sorting ability for the samples. Furthermore, an ideal AUC value combined with a worse ACC value denotes that many of the negative samples are classified as positive samples. That's because the negative and positive samples are unbalanced. For example, in the task of tank 1 against tank 2 and tank 3 classification, the samples of tank 1 are set as positive, and tank 2 and tank 3 are set as negative ones, which leads

to a sample unbalanced problem. However, when appropriate combination of λ_2 and λ_3 is chosen, the imbalance problem of samples will disappear.
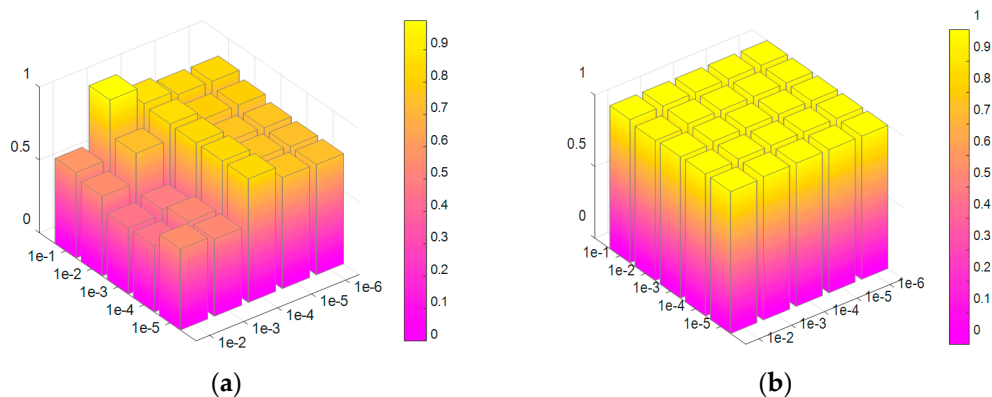
Figure 3. Average results of ACC and AUC versus parameter λ_2 and λ_3. (**a**) ACC results. (**b**) AUC results.

3.1.2. Comparison of Single Task and Multiple Task

To verify the effectiveness of multi-task learning, single task learning and multiple (three) tasks learning methods are compared, where parameter λ_1 is set as 1, λ_2 for 10^{-1} and λ_3 (c_3) for $\left\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\right\}$. The ACC results of three tanks are shown in Figure 4.
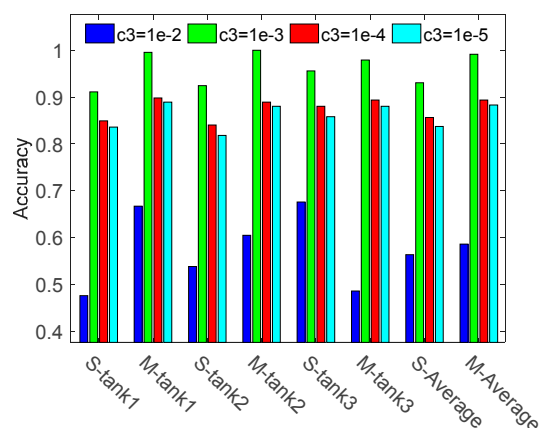
Figure 4. Recognition rate of three tanks when using different parameters λ_3 ($c3$) . The term 'S-tank1' ('M-tank1') denotes the recognition rate of tank1 in the framework of single (multiple) task learning, and 'S-Average' ('M-Average') means the average results of three tanks.

As shown in Figure 4, the overall recognition rate of three tasks jointly training method is 0.5856, 0.9915, 0.8933 and 0.8830 respectively when using different parameter λ_3. It is 2.26%, 6.11%, 3.70% and 4.59% better than that of the single task learning method. The result denotes that by jointly learning three tasks we can reveal and share the relationships among different tasks, which helps to discriminate the tanks with similar patterns. In our method, the relationships among different tasks can be automatically achieved by computing the learned covariance matrix **S**. The correlation coefficients of three tasks are shown in Figure 5, when λ_3 equals 10^{-3}.
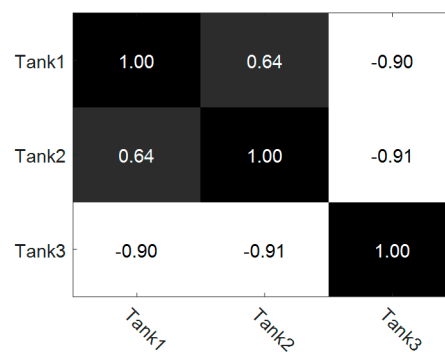
Figure 5. Correlation coefficients of three different tasks. The term 'Tank1' ('Tank2', 'Tank 3') denotes the task to classify tank 1 (tank 2, tank 3).

The relationship matrix shows that task 1 and task 2 have a high negative correlation with task 3. In the proposed three tasks jointly learning method, these relationships can be accurately described and utilized in the parameter space and projected space. Therefore, a 99.15% average recognition accuracy is achieved by this method. However, when these relationships are not properly handled, these strong correlations will degrade the model learning. Figure 4 shows that when λ_3 equals 10^{-2}, the ACC of tank 3 by multi-task learning is poor. The reason is that when a tight constraint is imposed on the projected space, the generalization ability of model will decline, which makes it difficult to accurately classify the highly correlated tank 3. Nevertheless, in most of the parameters λ_3, the performance of multi-task learning is better than that of single task learning.

3.1.3. Comparison against the State of the Art

To evaluate the performance of proposed method, our method is compared with the reference methods and the results are shown in Figure 6 and Table 3.
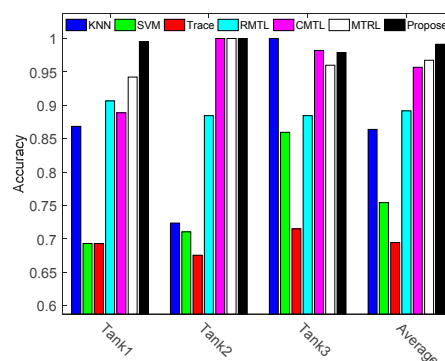
Figure 6. Recognition rates of all the compared methods. Term 'Average' denotes the average recognition rate of three tanks.

Table 3. Recognition rates of all the compared methods.

Method	KNN	SVM	Trace	RMTL	CMTL	MTRL	Proposed
Tank 1	0.8684	0.6929	0.6929	0.9066	0.9822	0.9600	0.9956
Tank 2	0.7237	0.7105	0.6754	0.8844	1.0000	1.0000	1.0000
Tank 3	1.0000	0.8596	0.7149	0.8844	0.8889	0.9522	0.9789
Average	0.8640	0.7543	0.6944	0.8918	0.9570	0.9674	0.9915

Figure 6 shows that the multi-task learning with a trace-norm regularization has the lowest recognition rate. It is because that the trace method learns a linear predictive function, which can't

accurately describe the nonlinear structures of HRRP data. This result suggests that it is necessary to extend the multi-task learning methods to nonlinear domains. Table 3 shows that the overall recognition rate for our method is 0.9915, compared to 0.9674 for MTRL, 0.9570 for CMTL, 0.8918 for RMTL, 0.6944 for Trace, 0.7543 for SVM and 0.8640 for KNN. It is 2.41%, 3.45%, 9.97%, 29.71%, 23.72% and 12.75% better than the competitors, MTRL, CMTL, RMTL, Trace, SVM and KNN, respectively. The simulation results denote that our method can accurately describe and utilize the three tasks relationships in the parameter space and projected space, which helps improve the recognition rate of radar targets with highly similar patterns.

3.2. Investigations Based on MSTAR Database

To further verify the effectiveness of the proposed method, extensive studies have been done based on the MSTAR public database, a gallery collected using a 10-GHz SAR sensor with 1 ft $\times$ 1 ft resolution in range and azimuth. Images are captured at various depressions over 0–359° range of aspect view. The sizes of the images are all around 128 $\times$ 128 pixels. To further avoid the influence of clutter, the images are cropped to 64 $\times$ 64 pixels. In this paper, the intensity of raw image is adopted as the feature. Specially, each raw image is concatenated into a 4096-dimensinal long vector. Then a 200-dimensional feature is achieved by PCA method. In the following numerical simulations, the parameters λ_1, λ_2 and λ_3 of our method are set as 1, 0.1 and 0.001, respectively.

3.2.1. Target Recognition under Standard Operating Conditions (SOC)

We first consider target recognition under SOC. Images acquired under operating condition of a 17° depression angle are used to train the mode, while the ones captured at an operating condition of a 15° depression angle are used for testing, as shown in Table 4. All ten targets are employed and their optical and SAR images are given in Figure 7. Among these vehicles, BMP2 and T72 have several variants with small structural modifications (denoted by series number). Only the standards (SN_9563 for BMP2 and SN_132 for T72) captured at 17° depression are available for training.
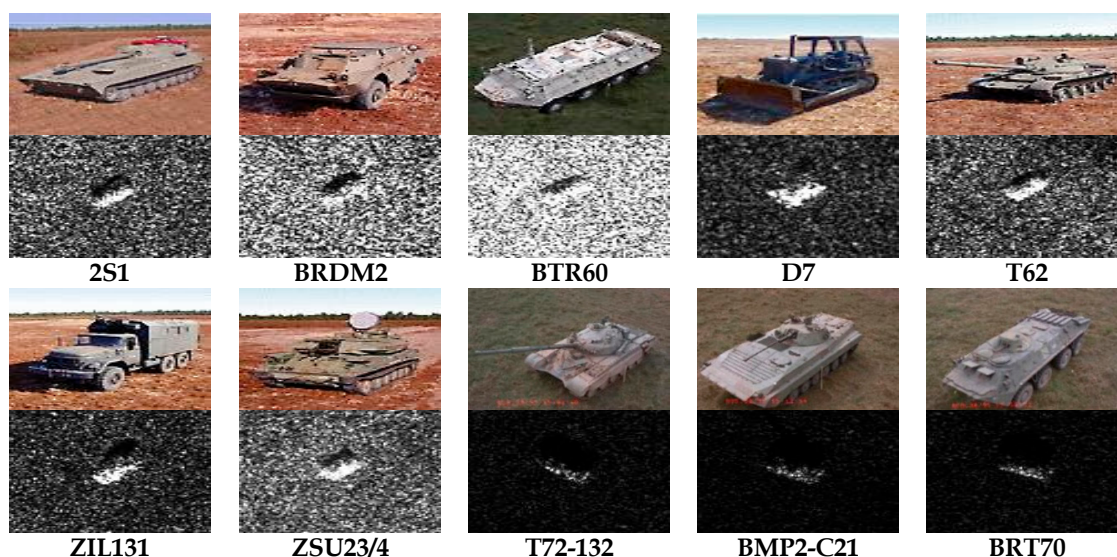

Figure 7. The optical and SAR images of ten targets to be recognized. The descriptions of these vehicles can be referred to [21].

Table 4. The number of images for training and testing about the ten targets under SOC.

Target	2S1	BRDM2	BTR60	D7	T62	ZIL131	ZSU23/4	BRT70	T72	BMP
Training (17°)	299	298	256	299	299	299	299	233	232(SN_132) 231(SN_812) 228(SN_s7)	233(SN_9563) 232(SN_9566) 233(SN_c21)
Testing (15°)	274	274	195	274	273	274	274	196	196(SN_132) 195(SN_812) 191(SN_s7)	195(SN_9563) 196(SN_9566) 196(SN_c21)

(A) Comparison of Single Task Learning and Multi-task Learning

In this section, single task learning method is compared with the other training method, like two tasks jointly learning, five tasks jointly learning, and ten tasks jointly learning method. Taking five tasks jointly learning method as an example, it is realized by diving the ten tasks into two equal groups and the tasks within the same groups are jointly learning. The formation of the other multi-task leaning method is similar to this one. The recognition results with different training methods are shown in Figure 8. Besides, the learned ten tasks relationships matrix is shown in Figure 9. As shown in Figure 8, the overall recognition rate of ten targets by ten modes jointly training is increased by 6.37%, 2.74%, and 2.14% respectively, compared with one mode individually training, two modes jointly training and five models jointly training. Moreover, ten modes jointly learning has a more robust ACC result compared with the other training methods. This is due to the fact that ten modes jointly learning method imposes a unified sparse constraint on all of the ten tasks and achieves a global balance in the process of training the ten models. Figure 9 shows that two groups of tasks ('2S1' and 'BMP2', 'T62' and'T72') are highly correlated. The single task leaning method can't appropriately handle and utilize these relationships, thus not being able to get a recognition rate as well as the multi-task leaning method.
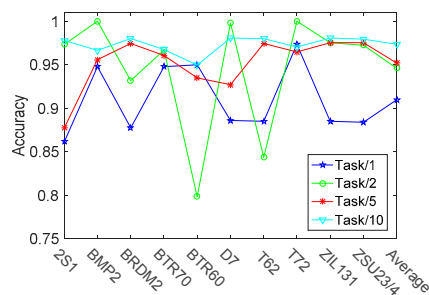
Figure 8. The recognition rates of ten vehicles with different learning methods. The term 'Task/1' denotes that each of the ten modes (classifiers) is individually learned (trained). While the term l 'Task/2' ('Task/5', 'Task/10') means that every two (five, ten) modes are jointly learned.
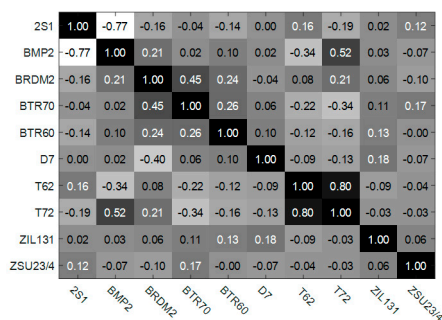
Figure 9. Correlation coefficients of ten different tasks. The term'2S1'means the task to recognize vehicle 2S1 and the meanings of other terms are similar to this one.

(B) Comparison against the State of the Art

Figure 10 and Table 5 show the comparison ACC results of the proposed method and the reference methods. From Table 5, we can see that the overall recognition rate of our method is 0.9734, compared to 0.9584 for MTRL, 0.9391 for CMTL, 0.9209 for RMTL, 0.7504 for Trace, 0.9017 for SVM and 0.9271 for KNN. It is 1.50%, 3.43%, 5.25%, 22.30%, 7.17% and 4.63% better than the competitors, MTRL, CMTL, RMTL, Trace, SVM and KNN, respectively. The results show that jointly training the ten modes under a unifying classification framework is beneficial for the improvement of recognition rate. Besides, Figure 10 also shows that the ACC of target '2S1' is lower than other targets in most of the reference methods. This may be because that, as shown in Figure 9, tasks '2S1' and 'BMP2' are highly correlated and most of the reference methods can't accurately describe and utilize this relationship, which results in a recognition performance degradation. On the contrary, the proposed method can utilize this relationship and get a better recognition accuracy.
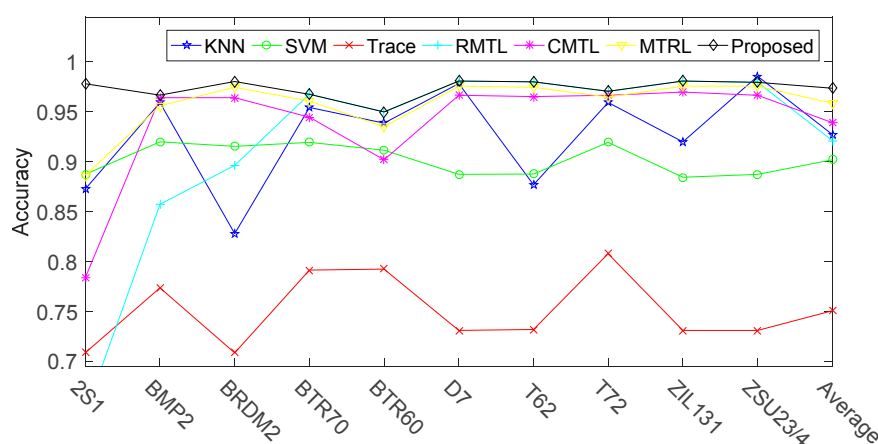

Figure 10. Recogniton rates of ten vehicles versus different methods.

Table 5. Recogniton rates of all the compared methods under SOC.

Methods	KNN	SVM	Trace	RMTL	CMTL	MTRL	Proposed
2S1	0.8723	0.8870	0.7082	0.6480	0.7833	0.8860	0.9780
BMP2	0.9590	0.9196	0.7733	0.8571	0.9641	0.9558	0.9665
BRDM2	0.8277	0.9151	0.7082	0.8960	0.9637	0.9757	0.9802
BRT70	0.9541	0.9192	0.7910	0.9674	0.9444	0.9606	0.9674
BTR60	0.9385	0.9113	0.7921	0.9497	0.9016	0.9350	0.9497
D7	0.9781	0.8870	0.7306	0.9806	0.9664	0.9754	0.9806
T62	0.8767	0.8874	0.7316	0.9799	0.9650	0.9731	0.9799
T72	0.9592	0.9192	0.8075	0.9703	0.9670	0.9646	0.9703
ZIL131	0.9197	0.8841	0.7412	0.9806	0.9696	0.9788	0.9806
ZSU23/4	0.9854	0.8870	0.7200	0.9794	0.9658	0.9720	0.9794
Average	0.9271	0.9017	0.7504	0.9209	0.9391	0.9584	0.9734

3.2.2. Target Recognition under Extended Operating Conditions (EOC)

To assure the practicability of our method, the recognition performances under different depression angles are assessed. Three vehicle targets 2S1, BRDM2, and ZSU23/4 are utilized. The images captured at an operating condition of a 17° depression are used to train the algorithm, while the ones collected at an operating condition of 30° and 45° depressions are used for testing, as shown in Table 6.

**Table 6.** The number of images for training and testing about the three targets under EOC.

Target	2S1	BRDM2	ZSU23/4
Training (17°)	299	298	299
Testing (30°)	288	287	288
Testing (45°)	303	303	303

(A) Comparison of Single Task Learning and Multi-task Learning

In this section, single task learning is compared with three tasks jointly learning. The ACC results and correlation coefficients matrix of three vehicles are shown in Figures 11 and 12 respectively.
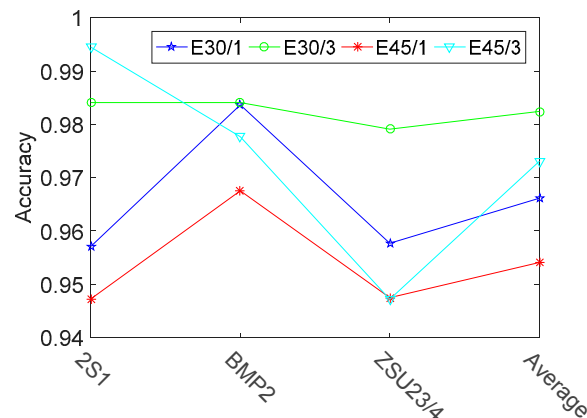
Figure 11. Recognition rates of three vehicles with different learning methods. The term 'E30/1' ('E45/1') denotes that each of the three modes is individually learned at an operating condition of a 17° depression and tested at an operating condition of 30° (45°) depression. Similar to the terms 'E30/1' and 'E45/1', the term 'E30/3' ('E45/3') denotes that all of the three modes are jointly learned.
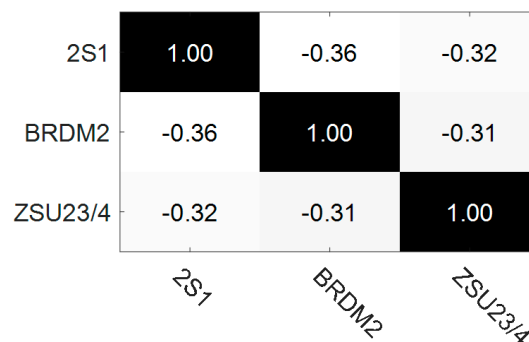

Figure 12. Correlation coefficients of three different tasks. The term '2S1' means the task to recognize vehicle 2S1 and the meanings of other terms are similar to this one.

Figure 11 indicates that the overall recognition rate of three modes jointly learning is 1.63% and 1.90% better than that of single mode individually learning under 30° and 45° testing depression angles, respectively. Figure 12 shows that the three tasks are connected with each other. The tasks connections in EOC test are not close enough, when compared with the relationships shown in Figure 5. One reason is that the SAR images contain a lot of speckle noises, while the simulated HRRP data does not contain noises. Nevertheless, the improvements of overall recognition rate are significant by means of three tasks jointly learning in both of the MSTAR data testing and HRRP data testing. All these results corroborate that multi-task learning is superior to single tasks learning.

(B) Comparison against the State of the Art

To evaluate the robustness of proposed method, the reference methods are compared with our method under two different depression angles. The ACC results are shown in Figure 13 and Table 7.

(**a**)

(**b**)
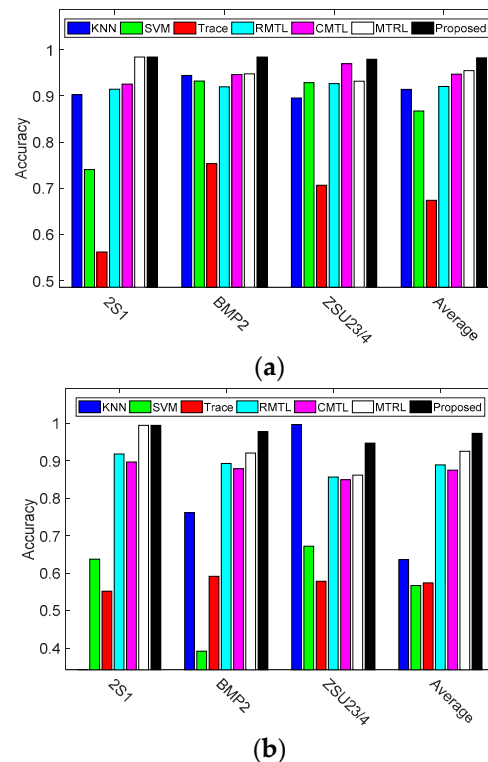
Figure 13. Recognition rates of three vehicles under EOC. (**a**) Testing under 30° depression; (**b**) Testing under 45° depression.

Table 7. Recognition rates of all the compared methods under EOC.

Methods	Training (17°)–Testing (30°)				Training (17°)–Testing (45°)			
	2S1	BMP2	ZSU23/4	Average	2S1	BMP2	ZSU23/4	Average
KNN	0.9028	0.9444	0.8955	0.9142	0.1505	0.7617	0.9967	0.6363
SVM	0.7409	0.9322	0.9288	0.8673	0.6373	0.3917	0.6719	0.5670
Trace	0.5625	0.7534	0.7067	0.6742	0.5518	0.5918	0.5787	0.5741
RMTL	0.9144	0.9199	0.9265	0.9203	0.9149	0.9195	0.8830	0.9058
CMTL	0.9254	0.9464	0.9697	0.9472	0.9166	0.8987	0.8692	0.8948
MTRL	0.9841	0.9477	0.9319	0.9546	0.9945	0.9481	0.9082	0.9502
Proposed	0.9841	0.9841	0.9791	0.9824	0.9945	0.9777	0.9472	0.9731

The overall recognition rate for the proposed method is 0.9824, compared to 0.9546 for MTRL, 0.9472 for CMTL, 0.9203 for RMTL, 0.6742 for Trace, 0.8673 for SVM and 0.9142 for KNN on the operating condition of 30° depression. It is 2.78%, 3.52%, 6.21%, 30.82%, 11.51% and 6.82% better than MTRL, CMTL, RMTL, Trace, SVM and KNN, respectively. When the algorithms are further evaluated using the images captured at an operating condition of 45°, the performances of all the methods are degraded, especially the KNN method. The recognition rate for KNN drops from 0.9142 to 0.6363. Different form KNN method, the multi-task leaning methods like RMTL, CMTL, MTRL and our method still keep a higher recognition rate due to the ability to share the multi-task relationships. Among all the approaches, the proposed method achieves the highest recognition rate of 0.9731. This is 2.29%, 7.83%, 6.73%, 39.90%, 40.61% and 33.68% better than MTRL, CMTL, RMTL, Trace,

SVM and KNN, respectively. The improvement of recognition accuracy is significant. The results demonstrate that the proposed method is much more robust toward depression variation than the reference methods.

From the above extensive analyses of simulation experiment results, we can draw the following conclusions. Multi-task relationships information can actually improve the performance of classification and automatically learning the task relationships from data is a more favorable option. Furthermore, a detailed and accurate description of the multi-task relationships in the original and projected space of model parameters is better than the rough hypothesis of multi-task relationships in terms of improving the radar target recognition performance.

4. Conclusions

In this paper, we propose a radar target recognition method based on the theory of clustered multi-task learning. In order to learn more useful and accurate relationships among multiple tasks, the potentially useful relationships in the projection space are further explored. The proposed method assumes that multi-tasks within the same cluster should be close to each other in the original and projected space, which contributes to discriminating the radar targets with similar patterns. Studies on the simulated HRRP data show that the proposed method can fully discover multi-task relationship in the projected space and accurately classify the targets with similar structures. Extensive comparative experiments on the MSTAR data are conducted to further demonstrate the superiority and robustness of the proposed method. The simulation experiment results under SOC and EOC demonstrate that the proposed method can properly reveal the latent relationships among multiple tasks and have a better performance than single task learning. Moreover, all of the comparisons against the state-of-the-art methods indicate the superiority of the proposed method.

Author Contributions: Luping Xu offered the SAR images and simulated HRRP data; Cong Li and Hua Zhang carried out the experiments and computer simulations and wrote the paper under the supervision of Weimin Bao.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pan, M.; Jiang, J.; Li, Z.; Cao, J.; Zhou, T. Radar HRRP recognition based on discriminant deep autoencoders with small training data size. *Electron. Lett.* **2016**, *52*, 1725–1727.
2. Zhou, D.Y. Radar target HRRP recognition based on reconstructive and discriminative dictionary learning. *Signal Process.* **2016**, *126*, 52–64. [CrossRef]
3. Sun, Y.G.; Du, L.; Wang, Y.; Wang, Y.H.; Hu, J. SAR automatic target recognition based on dictionary learning and joint dynamic sparse representation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *99*, 1777–1781. [CrossRef]
4. Wang, S.N.; Jiao, L.C.; Yang, S.Y.; Liu, H.Y. SAR image target recognition via complementary spatial pyramid coding. *Neurocomputing* **2016**, *196*, 125–132. [CrossRef]
5. Huang, X.Y.; Qiao, H.; Zhang, B. SAR target configuration recognition using tensor global and local discriminant embedding. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 222–226. [CrossRef]
6. Song, S.L.; Xu, B.; Yang, J. SAR target recognition via supervised discriminative dictionary learning and sparse representation of the SAR-HOG feature. *Remote Sens.* **2016**, *8*, 683. [CrossRef]
7. Kang, M.; Ji, K.; Leng, X.; Xing, X.; Zou, H. Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder. *Sensors* **2017**, *17*, 192. [CrossRef] [PubMed]
8. Zhao, Q.; Principe, J.C. Support vector machines for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 643–654. [CrossRef]
9. Sun, Y.J.; Liu, Z.P.; Todorovic, S.; Li, J. Adaptive boosting for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2007**, *43*, 112–125. [CrossRef]

10. Dong, G.G.; Kuang, G.Y.; Wang, N.; Zhao, L.J.; Lu, J. SAR target recognition via joint sparse representation of monogenic signal. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3316–3328. [CrossRef]

11. Zheng, H.; Geng, X.; Tao, D.C.; Jin, Z. A multi-task model for simultaneous face identification and facial expression recognition. *Neurocomputing* **2016**, *171*, 515–523. [CrossRef]

12. Liu, A.; Lu, Y.; Nie, W.Z.; Su, Y.T.; Yang, Z.X. HEp-2 cells classification via clustered multi-task learning. *Neurocomputing* **2016**, *195*, 195–201. [CrossRef]

13. Zhang, M.X.; Yang, Y.; Zhang, H.W.; Shen, F.M.; Zhang, D.X. L2, p-norm and sample constraint based feature selection and classification for AD diagnosis. *Neurocomputing* **2016**, *195*, 104–111. [CrossRef]

14. Zhang, Y.; Yeung, D.Y. A regularization approach to learning task relationships in multitask learning. *ACM Trans. Knowl. Discov. Data* **2014**, *8*, 1–31. [CrossRef]

15. Platt, J.C. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*; MSR-TR-98-14; Microsoft: Redmond, WA, USA, 1998.

16. Chen, X.; Pan, W.K.; Kwok, J.T.; Carbonell, J.G. Accelerated gradient method for multi-task sparse learning problem. In Proceedings of the Ninth IEEE International Conference on Data Mining, Miami, FL, USA, 6–9 December 2009; pp. 746–751.

17. Evgeniou, T.; Pontil, M. Regularized multi-task learning. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; pp. 109–117.

18. Zhou, J.Y.; Chen, J.H.; Ye, J.P. Clustered multi-task learning via alternating structure optimization. *Adv. Neural Inf. Process. Syst.* **2011**, *2011*, 702–710.

19. Zhou, Q.; Zhao, Q. Flexible clustered multi-task learning by learning representative tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 266–278. [CrossRef] [PubMed]

20. Guillaume, O.; Ben, T.; Michael, I.J. Joint covariate selection and joint subspace selection for multiple classification problems. *Stat. Comput.* **2010**, *20*, 231–252.

21. Song, H.; Ji, K.; Zhang, Y.; Xing, X.; Zou, H. Sparse representation-based SAR image target classification on the 10-class MSTAR data set. *Appl. Sci.* **2016**, *6*, 26. [CrossRef]