

Article

# Two-Stage Multi-Task Representation Learning for Synthetic Aperture Radar (SAR) Target Images Classification

Xinzheng Zhang <sup>1,\*</sup>, Yijian Wang <sup>1</sup>, Zhiying Tan <sup>1</sup>, Dong Li <sup>2</sup>, Shujun Liu <sup>1</sup>, Tao Wang <sup>2</sup> and Yongming Li <sup>1</sup>

<sup>1</sup> College of Communication Engineering, Chongqing University, Chongqing 400044, China; 20161202029t@cqu.edu.cn (Y.W.); 20161202031t@cqu.edu.cn (Z.T.); ly007@cqu.edu.cn (S.L.); yongmingli@cqu.edu.cn (Y.L.)

<sup>2</sup> Key Laboratory of Aircraft Tracking Telemetry & Command and Communication, Chongqing University, Chongqing 400044, China; dongli1983@cqu.edu.cn (D.L.); wt1977@cqu.edu.cn (T.W.)

\* Correspondence: zhangxinzheng@cqu.edu.cn; Tel.: +86-139-8322-7998

Received: 24 August 2017; Accepted: 28 October 2017; Published: 1 November 2017

**Abstract:** In this paper, we propose a two-stage multi-task learning representation method for the classification of synthetic aperture radar (SAR) target images. The first stage of the proposed approach uses multi-features joint sparse representation learning, modeled as a  $\ell_{2,1}$ -norm regularized multi-task sparse learning problem, to find an effective subset of training samples. Then, a new dictionary is constructed based on the training subset. The second stage of the method is to perform target images classification based on the new dictionary, utilizing multi-task collaborative representation. The proposed algorithm not only exploits the discrimination ability of multiple features but also greatly reduces the interference of atoms that are irrelevant to the test sample, thus effectively improving classification performance. Conducted with the Moving and Stationary Target Acquisition and Recognition (MSTAR) public SAR database, experimental results show that the proposed approach is effective and superior to many state-of-the-art methods.

**Keywords:** SAR; images classification; multitask learning; sparse representation; collaborative representation

## 1. Introduction

Synthetic aperture radar (SAR) are widely applied in various civil and military fields, such as aerial remote sensing to detect targets [1], environmental monitoring [2], and maritime surveillance [3]. Automatic target-recognition systems using SAR sensors continue to be developed for a number of applications, particularly in the area of military defense. The goal of these systems is to detect and classify military targets using various image- and signal-processing techniques. The conventional architecture of target-recognition systems consists of three separate stages: the pre-screener identifies local regions of interest using a constant false alarm rate (CFAR) detector, allowing all targets and numerous false alarms to pass. It is followed by a discriminator that aims to eliminate all natural false alarms. Finally, the classifier receives all man-made objects and attempts to categorize each input image as a specific target type contained in the training set [4]. General reviews of automatic target-recognition concepts and the SAR target-detection technologies can be found in [4,5]. This paper focuses on the final classification stage of the SAR automatic target-classification system. Target images obtained by SAR are significantly different with target optical images because microwave imaging is based on a scattering mechanism. Usually, SAR images are not as intuitive and fine as optical images. Moreover, the characteristics of SAR target images are very sensitive to azimuth and elevation

angles [4]. If these angles change significantly, the characteristics of the SAR target images also change significantly, which is challenging for their classification.

There are three mainstream paradigms governing SAR target images classification: template-based methods [5], model-based methods [6], and machine-learning based methods [7]. The template-based methods compare the test sample image with training sample images to determine the target category. These methods require a large number of target image templates and expensive computation. Moreover, background clutter also creates greater interference in template matching. For the model-based method, target images are described using a complex scattering model based on a scattering mechanism, and the target type is determined according to the likelihood estimation of the model parameters. However, since these complex scattering models are non-linear, it is difficult to obtain accurate estimation of the parameters. This method is easily affected by clutter interference and noise corruption. In addition, performance of the method at extended operating conditions will significantly decrease because of the obvious difference between training samples and test samples.

Recently, machine-learning based methods have attracted apparent interest for their use in SAR target images classification. One of the most popular methods is representation learning [8,9]. The most typical representation learning method is sparse representation. The sparse representation of one signal is based on the following theory: the observed signal can be represented by a linear combination of a series of known signals called atoms. Under sparse constraints ( $\ell_1$ -norm minimum) of the representative coefficient vector, a unique representative coefficient solution can be obtained, and the target type can be identified according to the minimum reconstruction error for each class. Sparse representation has been widely applied in a variety of recognition tasks, such as face recognition [10,11], speech recognition [12], and hyperspectral image classification [13]. For SAR target images classification, some researchers have investigated a number of approaches based on sparse representation. Zhang et al. proposed a method of multi-view joint sparse representation by exploring the correlation between target multi-view images [14]. Combining low-rank matrix recovery, Cheng et al. proposed an improved joint sparse representation approach to classify SAR target images [15]. Dong et al. investigated a method based on joint sparse representation with monogenic features, and developed the method on a Grassman manifold [16,17]. Furthermore, Liu et al. studied the Dempster–Shafer fusion of multiple sparse representations for target images recognition [18]. Sun et al. proposed a SAR target-recognition method combined with dynamic sparse representation and dictionary learning [19]. Song et al. reported an approach of supervised discriminant dictionary learning and sparse representation for SAR target recognition [20]. Lately, Liu et al. developed a new scatter center feature-extraction and target-recognition method based on sparse representation and a refinement dictionary [21].

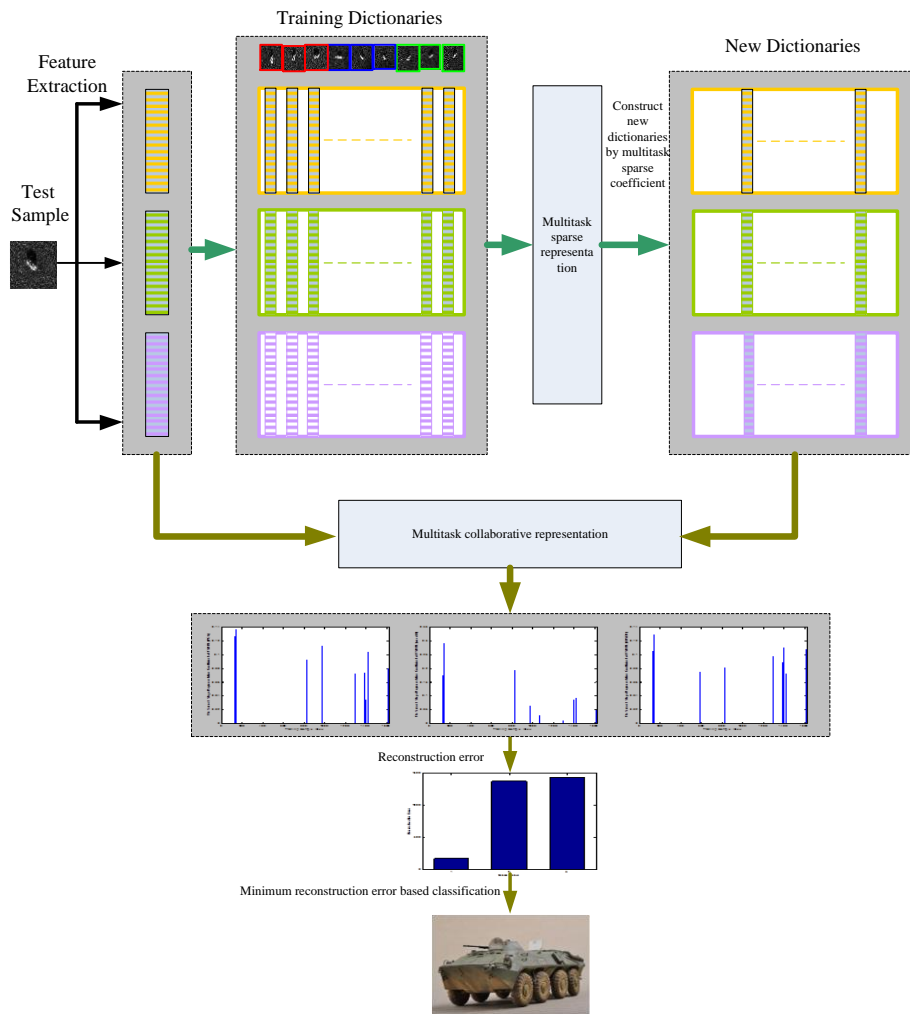
Another alternative method of representation learning is collaborative representation [22]. Collaborative representation can be interpreted as an optimization problem under  $\ell_2$ -norm minimum constraints. It has also been widely used in face recognition, hyperspectral classification, etc. [23]. Compared to sparse representation, collaborative representation can greatly reduce the computational cost while maintaining high classification accuracy [23].

In recent years, the combination of multi-task learning and representation learning has become an important trend in pattern recognition. The idea behind multi-task learning is that when these tasks are similar enough or correlative to a certain extent, exploiting the correlation between them is beneficial for improving the generalization of recognition. Some research results have confirmed the advantages of multi-task learning [24]. Yan et al. extracted various types of features from images, and viewed each type's sparse representation recognition as a task in order to model multi-feature joint sparse representation recognition as a multi-task sparse learning problem [25]. Fang et al. proposed a face-recognition method based on local Gabor feature adaptive multi-task sparse representation [26]. Li et al. proposed a hyperspectral image-classification method by combining multi-task learning with collaborative representation [27]. Luo et al. investigated a manifold regularized multi-task learning (MRMTL) algorithm, which can effectively control the model complexity and ensure that the functions

in the shared hypothesis space are smooth along the data manifold [28]. Furthermore, Luo et al. proposed a novel large margin multi-modal multi-task feature extraction framework, which can not only handle correlated and noisy features but also utilize the complementarity of different modalities to reduce feature redundancy [29]. Both multi-task sparse representation and multi-task collaborative representation can exploit shared training sample patterns between different tasks to ensure that the correct training samples are selected, excluding interferential atoms. In addition, recently some SAR target-recognition technologies have been based on deep learning. Ding et al. developed the target-recognition algorithm combined with a data enhancement convolutional neural network [30]. Chen et al. investigated the application of deep convolution networks for SAR target recognition in detail [31].

In this work, we propose a two stage multi-task representation learning method for SAR target images classification. Figure 1 shows a schematic view of the approach. The approach first extracts three kinds of features for all training samples and test samples, which include principal component analysis (PCA) features, wavelet transform features, and 2D Slice Zernike Moments (2DSZM) features [32]. The first stage is to represent each feature of the test sample as a linear combination of the corresponding features of the training set, and to determine the  $Q$  neighbor samples of the test sample in the training set by using multi-task sparse representation. This is because, in principle, the current test sample and its neighboring samples should come from the same class, which means that  $Q$  neighboring samples make the greatest contribution to identifying the test sample. Thus, the first step in this algorithm is to detect training samples that are far from the test sample, it being assumed that these samples have no effect on the classification decision. This is helpful for accurate classification of test samples. In fact, using part of the training samples rather than all training samples to identify a test sample can greatly reduce the interference of those training samples that are far away from the test sample. The second stage of this method is to represent the test sample with the new dictionary consisting of  $Q$  neighbors in the framework of multi-task collaborative representation. Furthermore, the representation results are used to infer the test sample label. We adopt multi-task collaborative representation in the second stage because of its uncomplicated closed solution and low computation cost. The proposed method is based on the following factors: the first stage confirms a number of training samples that are most relevant to the current test sample. Since the class labels of the chosen training samples are usually a subset of all class labels, the final classification becomes a problem that determines the test sample label from a small number of candidate class labels. It would be beneficial to have an accurate inference in the second stage that the real label of the test sample was one of the training subset labels. The proposed approach not only exploits the ability of combined multiple features representation learning but also greatly reduces the interference of those irrelevant atoms in the dictionary, which leads to enhanced classification performance. Evaluation of the proposed method is conducted with the Moving and Stationary Target Acquisition and Recognition (MSTAR) benchmark data sets. Experimental results validate the effectiveness and superiority of the proposed approach.

The paper is organized as follows. In Section 2, we briefly describe the three types of feature-extraction methods used in the work, and review basic sparse representation and collaborative representation. In Section 3, a two-stage multi-task representation learning algorithm is developed in detail. In Section 4, experiments are carried out with the MSTAR database, and the performance of the proposed approach is described. Finally, we conclude the paper in Section 5.



**Figure 1.** Flowchart of two-stage multi-task representation learning for synthetic aperture radar (SAR) target images classification.

## 2. Multiple Features Extraction and Representation Learning Classifier

In this section, we introduce the three types of feature extraction adopted in this paper, and present the sparse representation classifier and collaborative representation classifier.

### 2.1. Multiple Features Extraction

It is well known that each type of feature describes the image from different aspects, and a single feature cannot contain all discriminative information. In addition, it is impractical to extract an optimal feature. For image-classification systems, a better approach is to combine multiple features rather than use a single feature [33]. In this paper, three types of features were extracted from SAR target images, including PCA features, wavelet transform features, and 2DSZM features.

#### 2.1.1. PCA Features Extraction

PCA is a mature feature-extraction technique. The aim of PCA is to find a projection vector, that can represent the original sample in the greatest degree and map the original sample to a low-dimensional space through a linear transformation. Let  $\mathbf{y} \in \mathbb{R}^{1 \times N}$  denotes the test sample, and  $\mathbf{P} \in \mathbb{R}^{N \times r}$  denotes the projection matrix; then, the PCA features extraction can be expressed as follows:

$$\mathbf{s}_{\text{PCA}} = \mathbf{y}\mathbf{P} \quad (1)$$

where  $s_{PCA}$  is the extracted PCA features.

### 2.1.2. Wavelet Transform Feature Extraction

Wavelet transform is widely applied in image feature extraction, image compression, and so on. In this work, we apply two-dimensional discrete wavelet transform, and vectorize the first-level low-frequency sub-band images as feature vectors. Wavelet feature extraction can be expressed as follows:

$$s_{WAVELET} = DWT2(y) \quad (2)$$

where  $DWT2(\bullet)$  represents the two-dimensional discrete wavelet transform, and  $s_{WAVELET}$  donates the extracted wavelet transform feature vector.

### 2.1.3. 2DSZM Feature Extraction

A novel feature named 2DSZM was developed recently, depicting SAR target images scattering features effectively [32]. SAR target images are usually expressed as gray images. According to the electromagnetic scattering theory, each pixel value of a microwave image is related to the intensity of the backscattered wave. In SAR target images, there are strong scattering centers, moderate scattering centers, and weak scattering centers. Since the location and intensity distribution of different scattering regions are not same, different scattering regions can be separated by the slices operation. This operation is called the 2D Slice operation. We extract the Zernike moments of each slice, and then connect all slices' Zernike moments to get the 2DSZM feature vector donated by  $s_{2DSZM}$ . The Zernike moments can be obtained by Equation (3). Details of the 2DSZM feature extraction can be found in the literature [32].

$$Z_{pq} = \frac{p+1}{\pi} \iint_{x^2+y^2 \leq 1} V_{pq}^*(x,y) f(x,y) dx dy \quad (3)$$

## 2.2. Sparse Representation Classifier

The success of the sparse signal model can be attributed to the sparse representation of the observed signal with a set of known observation sets (dictionaries). The optimal sparse representation can be obtained effectively by convex optimization methods [34]. We assume that there are  $k$  ( $k = 1, \dots, K$ ) classes of training data, denoted by  $X_k = [x_{k,1}, x_{k,2}, \dots, x_{k,n_k}] \in \mathbb{R}^{m \times n_k}$ , where the samples are arranged in  $\mathbb{R}^m$  with the form of column vectors. The unknown observed sample  $y \in \mathbb{R}^m$  is represented as a linear combination of the  $k$ -th class training samples with the representation coefficient  $\alpha_k = [\alpha_{k,1}, \alpha_{k,2}, \dots, \alpha_{k,n_k}]^T \in \mathbb{R}^{n_k}$ .

$$y = x_{k,1}\alpha_{k,1} + x_{k,2}\alpha_{k,2} + \dots + x_{k,n_k}\alpha_{k,n_k} \quad (4)$$

Because the class label of the observed sample  $y$  is unknown, it is necessary to use all the  $k$  classes training samples to represent it. Let  $X = [X_1, X_2, \dots, X_K] \in \mathbb{R}^{m \times n}$ , where  $n = \sum_{k=1}^K n_k$  is the total number of training samples. Then, the linear representation of  $y$  can be written as follows:

$$y = X_1\alpha_1 + X_2\alpha_2 + \dots + X_K\alpha_K = X\alpha \quad (5)$$

where  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_K]^T \in \mathbb{R}^n$  is the representation coefficient vector. Theoretically, the elements in the vector  $\alpha$  corresponding to atoms with the same class as  $y$  are non-zero, and other elements are zero. The general method is to find the sparsest representation by adding a regularization term to the representation-coefficient vector [35].

$$\min_{\alpha} \|\alpha\|_0 \text{ s.t. } \|y - X\alpha\|_2 \leq \varepsilon \quad (6)$$

where  $\|\cdot\|_0$  represents the  $\ell_0$ -norm of  $\alpha$ , and  $\varepsilon$  donates tolerance error. The optimization of the  $\ell_0$ -norm minimum problem is NP-hard, which is usually relaxed to a  $\ell_1$ -norm minimum problem as follows:

$$\min_{\alpha} \|\mathbf{y} - \mathbf{X}\alpha\|_2^2 + \eta \|\alpha\|_1 \quad (7)$$

where  $\eta$  is the balance parameter, and  $\|\cdot\|_1$  donates  $\ell_1$ -norm of  $\alpha$ . Since Equation (7) is a convex optimization problem, it can be solved by conventional optimization algorithms [36].

After obtaining the optimal sparse representation vector  $\hat{\alpha}$ , the label of  $\mathbf{y}$  can be obtained by the minimum reconstruction error discriminant criterion as follows:

$$\text{identity}(\mathbf{y}) = \min_{k=1,\dots,K} \|\mathbf{y} - \mathbf{X}_k \hat{\alpha}\|_2 \quad (8)$$

The aforementioned outcome is the famous sparse representation classifier (SRC) [11].

### 2.3. Collaborative Representation Classifier

As an alternative representation-learning theory, the collaborative representation classifier (CRC), has been developed more fully recently [37]. Collaborative representation has been widely used in face recognition, hyperspectral classification, etc. A difference with SRC is that CRC is essentially an  $\ell_2$ -norm optimal problem. The collaborative representation of the observed sample  $\mathbf{y}$  under the training set  $\mathbf{X}$  can be written as follows:

$$\hat{\beta} = \arg \min_{\beta} \{\|\mathbf{y} - \mathbf{X}\beta\|_2 + \gamma \|\beta\|_2^2\} \quad (9)$$

where  $\gamma$  is the balance parameter,  $\beta$  is the representation coefficient vector, and  $\|\cdot\|_2$  donates the  $\ell_2$ -norm of  $\beta$ .

Compared to SRC, the solving process of CRC needs no optimization search; therefore, the computation cost is much lower than SRC. Equation (9) of CRC has a closed solution as follows:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X} + \gamma \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}. \quad (10)$$

With  $\hat{\beta}$ , the class label of  $\mathbf{y}$  can be decided by the minimum reconstruction error criterion as follows:

$$\text{identity}(\mathbf{y}) = \min_{k=1,\dots,K} \|\mathbf{y} - \mathbf{X}_k \hat{\beta}\|_2. \quad (11)$$

## 3. Two-Stage Multi-Task Representation Learning

For image-classification tasks, taking into account the shared patterns among various tasks is beneficial for improving the generalization performance of classification. Numerous studies have confirmed the superior performance of the multi-task learning framework in theory and practice. For example, multi-task sparse representation has been applied in face recognition, and multi-task collaborative representation has been used in hyperspectral image classification. In this work, the classification by each type of feature can be considered as an individual task. Multi-feature classification constitutes a multi-task model, in which multiple tasks should share feature subsets of the training set for a test sample.

Both multi-task sparse representation and multi-task collaborative representation utilize all training samples for classification. However, many studies show that the local subset rather than all of the training samples mainly contributes to classification. That is, many training samples play the role of interference. For instance, it has been illustrated that local PCA is superior to global PCA in classification [38]. Vural proposed the use of local dependencies of samples to conduct the classification [39], while Xu et al. investigated two-phase sparse representation with a local subset of training samples [40].



In this section, we propose a two-stage multi-task representation learning (TSMRL) algorithm for SAR target images classification. The first stage of the algorithm is to represent each feature of the test sample as a linear combination of the corresponding features of all training samples. The  $\ell_{2,1}$ -norm regularized multi-task sparse learning is adopted to determine the  $Q$  nearest neighbor training samples for the test sample. Using local training samples instead of all training samples to identify a test sample can greatly reduce the interference of those training samples far away from the test sample. The new dictionary is constructed by the  $Q$  nearest neighbor training samples. In the second stage of the algorithm, multi-task collaborative representation is used for classification with the new dictionary, leading to the final decision of the test sample. In this section, we will describe the proposed two-stage multi-task representation learning method in detail.

### 3.1. The First Stage: $\ell_{2,1}$ -Norm Regularized Multi-Task Sparse Representation Learning

Firstly, it is assumed that the test sample features and the training samples features approximately satisfy the following equation:

$$\mathbf{y}^k = a_{11}^k \mathbf{x}_{11}^k + a_{12}^k \mathbf{x}_{12}^k + \dots + a_{1n_1}^k \mathbf{x}_{1n_1}^k + \dots + a_{C1}^k \mathbf{x}_{C1}^k + \dots + a_{Cn_C}^k \mathbf{x}_{Cn_C}^k, \quad (12)$$

where  $\mathbf{y}^k (k = 1, 2, 3)$  here is the  $k$ -th pattern features of the test sample, and  $a_{ci}^k (c = 1, 2, \dots, C; i = 1, 2, \dots, n_c; k = 1, 2, 3)$  is the representation coefficient. Equation (12) can be rewritten in the following matrix form:

$$\mathbf{y}^k = \mathbf{X}^k \mathbf{a}^k. \quad (13)$$

According to the sparse representation principle, each pattern feature of the test sample should share features from the same training samples. Taking into account errors caused by noise, we can achieve the following  $\ell_{2,1}$ -norm regularized multi-task sparse representation model:

$$\arg \min_A \left\{ \sum_{k=1}^3 \|\mathbf{y}^k - \mathbf{X}^k \mathbf{a}^k\|_F^2 + \lambda \|\mathbf{A}\|_{2,1} \right\}, \quad (14)$$

where  $\lambda$  is the balance parameter,  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix, and  $\|\cdot\|_{2,1}$  denotes the  $\ell_{2,1}$ -norm,  $\mathbf{A} = [\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3]$ .

The regularization term in Equation (14) is chosen as the  $\ell_{2,1}$ -norm regularization term of  $\mathbf{A}$ . This is because, for all kinds of features from the same test sample, the location of non-zero coefficients in the corresponding sparse coefficient vectors should be similar. Additionally, the corresponding coefficient values of these shared atoms are different due to differences between feature types. Under this assumption, non-zero coefficients of the representation coefficient matrix should be in the same row. A regularized  $\ell_{2,1}$ -norm can be imposed on  $\mathbf{A}$  to select a small number of non-zero rows. The optimal problem (14) can be obtained by the accelerated proximal gradient algorithm [41].

As is known, each column of  $\mathbf{A}$  is a corresponding representation coefficient vector of each feature. For representing the test sample, each training sample contributes to classification differently. The contribution of a training sample can be estimated by the corresponding representation coefficient value. A large coefficient value means that the training sample makes a large contribution to the representation. Since there are three representation coefficient vectors, we adopt the following method to obtain a local training subset comprising  $Q$  nearest neighbor samples. First, the  $Q$  atoms with the largest coefficient values are selected in each coefficient vector, leading to three groups. Due to the  $\ell_{2,1}$ -norm regularized multi-task sparse constraints, most of the  $Q$  atoms come from the same training samples. However, there are still a few atoms from different training samples. A new subset will be formed by three groups. We get a subset of  $V$  atoms after mixing all three groups and merging the same atoms. We assume that  $\mathbf{x}_j^k (j = 1, \dots, V)$  is the  $k$ -th pattern feature vector of the  $j$ -th atom in the  $V$  subset, and let  $e_j = \sum_{k=1}^3 \|\mathbf{y}^k - \mathbf{x}_j^k\|_2^2$ . According to the ascending order of  $e_j$ , we select the first  $Q$  atoms

from  $V$  to form the final nearest neighbor local subset. For each type of feature, the final  $Q$  nearest neighbor atoms are reserved, while other atoms are set to zero vectors. In this way, the dictionary  $\mathbf{X}^k$  is updated to a new dictionary  $\mathbf{X}'^k$ . Thus, the interference caused by irrelevant atoms can be greatly reduced, which can improve correct judgement.

### 3.2. The Second Stage: Multi-Task Collaborative Representation Learning

The second stage of the TSMRL algorithm is to represent the test sample based on the multi-task collaborative representation with the new dictionary  $\mathbf{X}'^k$ , as follows:

$$\operatorname{argmin}_{\mathbf{B}} \left\{ \sum_{k=1}^3 \|\mathbf{y}^k - \mathbf{X}'^k \mathbf{b}^k\|_F^2 + \rho \|\mathbf{B}\|_F^2 \right\}, \quad (15)$$

where  $\rho$  is the balance parameter.  $\mathbf{B}$  is the collaborative representation coefficient matrix, and  $\mathbf{B} = [\mathbf{b}^1, \mathbf{b}^2, \mathbf{b}^3]$ . The optimal problem of multi-task collaborative representation has analytical solution as follows [27]:

$$\hat{\mathbf{b}}^k = \left( (\mathbf{X}'^k)^T \mathbf{X}'^k + \rho \mathbf{I} \right)^{-1} (\mathbf{X}'^k)^T \mathbf{y}^k. \quad (16)$$

We employ multi-task collaborative representation because of its simplicity in computation and its accurate classification ability.

Then, the class label of  $\mathbf{y}$  is predicted to the class with the lowest total reconstruction error accumulated over all tasks, i.e.,

$$\operatorname{Class}(\mathbf{y}) = \operatorname{argmin}_{c \in \{1, 2, \dots, C\}} \sum_{k=1}^3 \|\mathbf{y}^k - \mathbf{X}'^k \hat{\mathbf{b}}_c^k\|_2^2. \quad (17)$$

In summary, the steps of TSMRL are shown in Algorithm 1.

---

#### Algorithm 1 Two-stage multi-task representation learning for SAR target images classification

---

Input:

$\mathbf{X}_{m \times n}$ : All training samples

$\mathbf{Y}_{m \times p}$ : All test samples

Output: the identity of  $\mathbf{Y}_{m \times p}$

Steps:

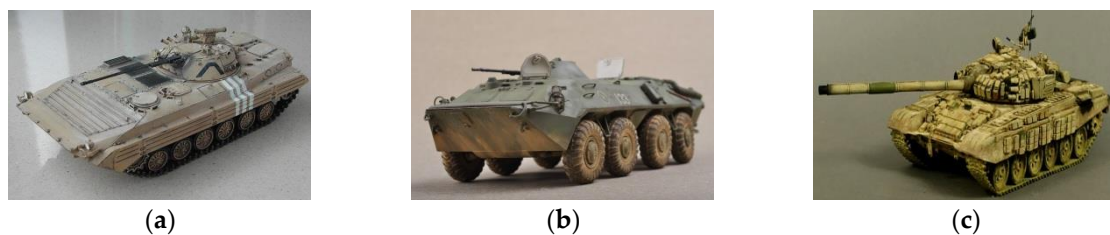
- 1) Extract three types of features from  $\mathbf{X}_{m \times n}$  respectively; let  $\mathbf{X}^1 = \mathbf{X}_{\text{PCA}}$ ,  $\mathbf{X}^2 = \mathbf{X}_{\text{WAVELET}}$ , and  $\mathbf{X}^3 = \mathbf{X}_{\text{2DSZM}}$ .
  - 2) Select a test sample  $\mathbf{y}_i$  from  $\mathbf{Y}_{m \times p}$ , and three types of features of  $\mathbf{y}_i$  are extracted; let  $\mathbf{y}_i^1 = \mathbf{y}_{i\text{PCA}}$ ,  $\mathbf{y}_i^2 = \mathbf{y}_{i\text{WAVELET}}$ , and  $\mathbf{y}_i^3 = \mathbf{y}_{i\text{2DSZM}}$ .
  - 3) Using  $\ell_{2,1}$ -norm regularized multi-task sparse representation, represent  $\{\mathbf{y}_i^1, \mathbf{y}_i^2, \mathbf{y}_i^3\}$  with the dictionaries  $\{\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3\}$ ; obtain the representation matrix  $\mathbf{A}$ . Then, obtain the local subset comprising  $Q$  nearest neighbors, and construct new dictionaries  $\{\mathbf{X}'^1, \mathbf{X}'^2, \mathbf{X}'^3\}$ .
  - 4) Using multi-task collaborative representation, represent  $\{\mathbf{y}_i^1, \mathbf{y}_i^2, \mathbf{y}_i^3\}$  with the new dictionaries  $\{\mathbf{X}'^1, \mathbf{X}'^2, \mathbf{X}'^3\}$ . Then, obtain the representative matrix  $\mathbf{B}$ .
  - 5) Decide the label of the test sample  $\mathbf{y}_i$  based on the criterion of the total minimum reconstruction error of multi-task collaborative representation.
  - 6) If all testing samples are classified, go to step 7). Otherwise, return to step 1).
  - 7) End
-



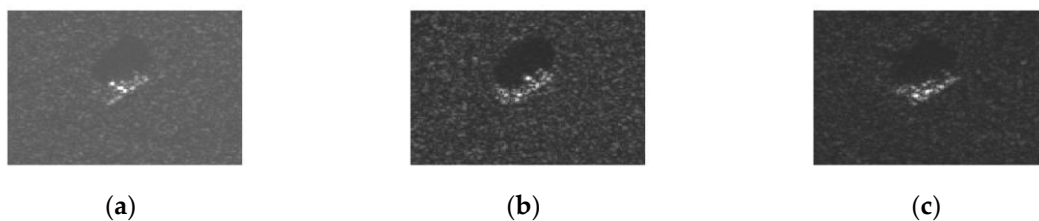
## 4. Experimental Results

### 4.1. Experimental Data Set

In order to evaluate the performance of the proposed approach, we used the MSTAR SAR image database to carry out experiments. SAR is a typical microwave imaging sensor that can obtain high-resolution two-dimensional images of targets. The database is widely used for the performance evaluation of various SAR target-recognition algorithms. The database was acquired by X band SAR sensors with  $0.3 \text{ m} \times 0.3 \text{ m}$  resolutions. The main targets in the database are armored vehicles, tanks, and other vehicles. SAR images in the database have an azimuth range from  $0^\circ$ – $360^\circ$ , and have two different depression angles ( $15^\circ$ ,  $17^\circ$ ). In the experiments, we mainly used images of three targets, namely BMP2, BTR70 and T72. The optical images of the three targets are given in Figure 2. The SAR images of the three targets with similar azimuth angles are given in Figure 3. In particular, BMP2 has three different configurations: BMP2-9563, BMP2-9566, and BMP2-C21. BTR70 has only one configuration, BTR70-C71. T72 also has three different configurations: T72-132, T72-812 and T72-S7. Table 1 lists the number of training samples and test samples. SAR images of the three targets obtained at a  $17^\circ$  depression angle were used as training samples, while images acquired at a  $15^\circ$  depression angle were used as test samples. In addition, a SAR image from the MSTAR database contains background clutter besides the target; and the target is usually located in the central region of the SAR image. In order to reduce the interference of background clutter, we used a  $60 \times 60$  window to get the central region of the original SAR image.



**Figure 2.** Optical images of three targets. (a) BMP2; (b) BTR70; (c) T72.



**Figure 3.** SAR images of three targets. (a) BMP2 ( $45.5^\circ$ ); (b) BTR70 ( $45.0^\circ$ ); (c) T72 ( $44.8^\circ$ ).

**Table 1.** The targets and number of training and test datasets.

Number of Objects	1	2	3	4	5	6	7
Training sample type ( $17^\circ$ )	BMP2 sn-9563	BMP2 sn-9566	BMP2 sn-c21	BTR70 sn-c71	T72 sn-132	T72 sn-812	T72 sn-s7
Number	233	232	233	233	232	231	228
Testing sample type ( $15^\circ$ )	BMP2 sn-9563	BMP2 sn-9566	BMP2 sn-c21	BTR70 sn-c71	T72 sn-132	T72 sn-812	T72 sn-s7
Number	195	196	196	196	196	195	191

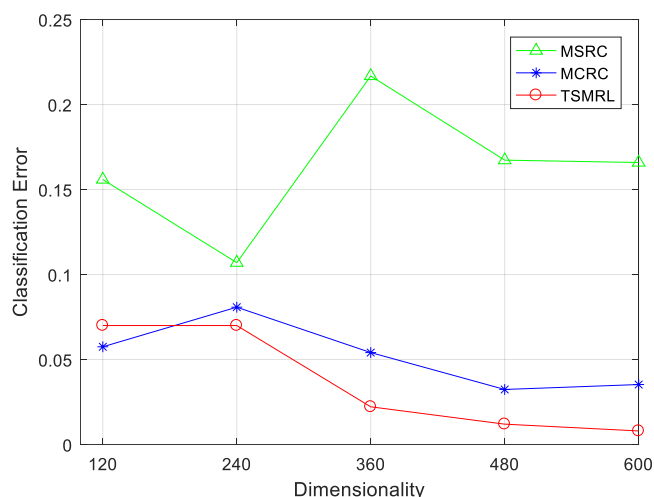
### 4.2. Classification Results and Analysis

We first analyzed the classification performance of the proposed TSMRL algorithm. In order to illustrate the effectiveness of the TSMRL, we compared it to the multi-task sparse representation classifier (MSRC) and the multi-task collaborative representation classifier (MCRC). The advantage

of TSMRL lies in the fact that the best local subset for the test sample can be obtained through the  $\ell_{2,1}$ -norm regularized multi-task joint sparse in the first stage, which can greatly reduce the interference of irrelevant atoms. In fact, this agrees with the characteristics of SAR target images, which are sensitive to azimuth changes. Thus, only training samples from the same target and with adjacent azimuth can be selected as nearest neighbors of the test sample.

#### 4.2.1. Classification Performance under Different Feature Dimensions

First, we compared the classification performance of the three methods under different feature dimensions. The results are shown in Figure 4 and Table 2. It is necessary to note that the dimensionality of three features used in this paper is different. Therefore, the change of feature dimensions was obtained by decreasing each feature dimension with the same ratio. The horizontal axis in Figure 4 is the dimensionality of the PCA feature. As can be seen from Figure 4, the error rate of TSMRL decreases with the increase of feature dimension. When the dimension is 600, TSMRL had the lowest error rate of 0.80%, that is, the average classification accuracy was 99.20%. Under the same feature dimension, the error rates of MCRC and MSRC were 3.53% and 16.56%, respectively, which are much higher than that of the TSMRL. Moreover, the TSMRL can maintain small error rates in lower dimensions. These experimental results support the advantages of the proposed algorithm for SAR target images classification.



**Figure 4.** Classification performance of three algorithms with varying feature dimensions.

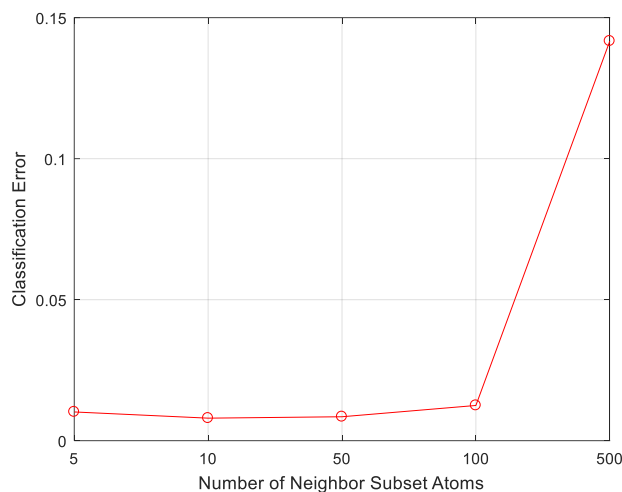
**Table 2.** Error rates of the three algorithms with varying feature dimensions.

Dimensionality	Method		
	TSMRL	MSRC	MCRC
120	7.00%	15.59%	5.75%
240	7.00%	10.70%	8.08%
360	2.22%	21.68%	5.41%
480	1.20%	16.73%	3.24%
600	0.80%	16.56%	3.53%

#### 4.2.2. Classification Performance under Different Number of Neighbor Atoms

The dictionaries in both MSRC and MCRC are composed of all training samples. However, in TSMRL, a subset of training samples is selected after the first stage  $\ell_{2,1}$ -norm regularized multi-task joint sparse learning. Thus, although the dictionary size is the same as those of MSRC and MCRC, the number of valid non-zero atoms is much smaller than those of MSRC and MCRC.

We changed the value of  $Q$  in order to observe the impact of the number of selected neighbor atoms. The classification results are shown in Figure 5.



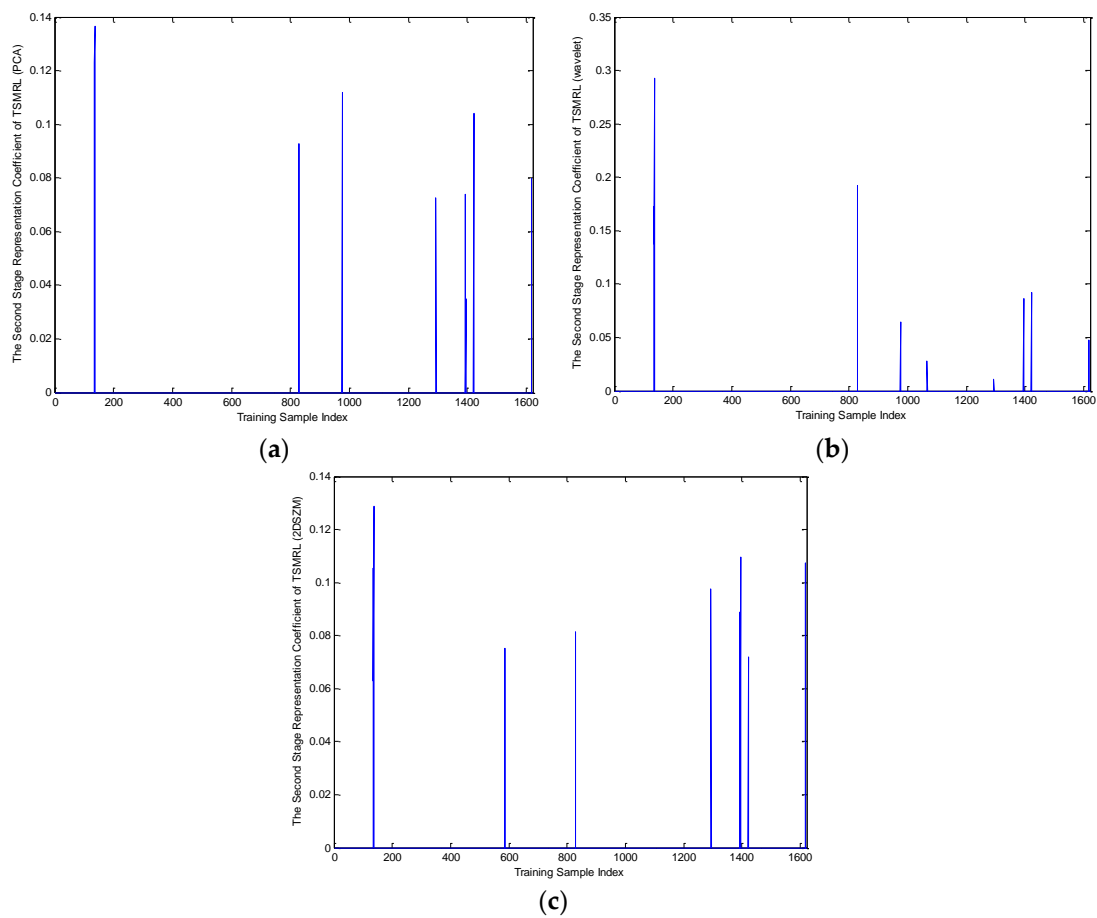
**Figure 5.** Classification performance of two-stage multi-task representation learning (TSMRL) under different number of neighbor subset atoms.

As shown in Figure 5, when  $Q$  equals 10, the error rate of TSMRL attains the minimum of 0.80%. With the increased  $Q$ , the error rate becomes larger, and reaches 14.17% when  $Q$  equals 500. This is mainly because many interfering atoms are inevitably introduced with the increased number of neighbor atoms. Nevertheless, it can be seen from Figure 5 that TSMRL still maintains low error rates over a wide range of  $Q$  values from 5 to 100. According to the above results, the number of neighbor atoms was set to 10 in subsequent experiments.

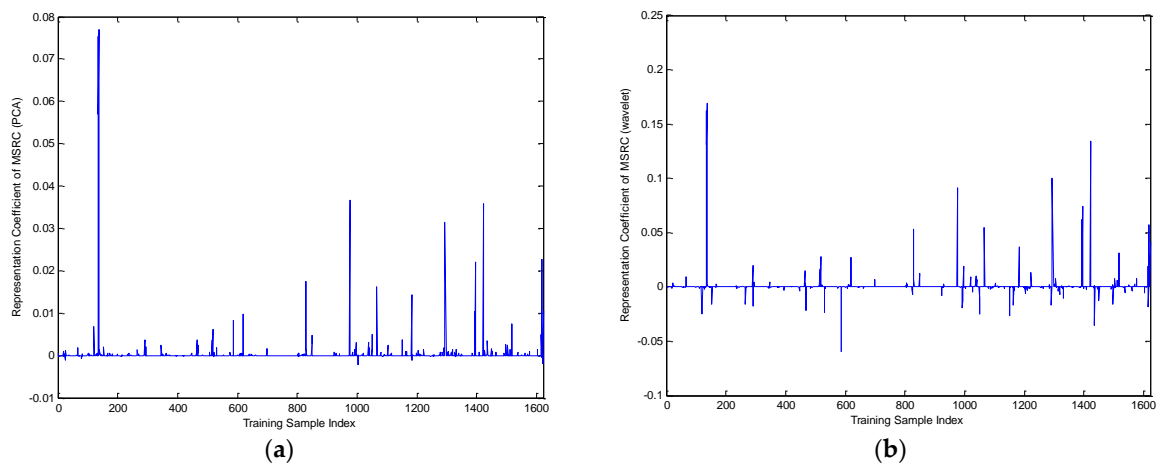
To compare the performance of the three methods definitively, one query sample of BMP2-9563 with an azimuth angle of  $172.5^\circ$  was selected to analyze the representation coefficient and reconstruction error, as shown in Figures 6–10. Compared to the representation coefficient of both MSRC and MCRC, the representative coefficient vector of TSMRL becomes much sparser, as shown in Figure 7. The reason for this is that we selected a fixed number of neighbor atoms ( $Q = 10$ ) for TSMRL. However, for both MSRC and MCRC, there are many atoms with large coefficients belonging to non-BMP2 in the representation coefficient vector, which causes huge interference. The reconstruction error graphs of the three methods shown in Figure 10 confirm our predictions. Figure 10 suggests that minimal reconstruction errors for both MSRC and MCRC correspond to the incorrect class, while the minimal reconstruction error for TSMRL means that it is able to correctly identify the target. As shown in Figure 10a, the reconstruction error of the query sample in TSMRL is far less than other target categories. In summary, TSMRL can identify targets more accurately than MSRC and MCRC.



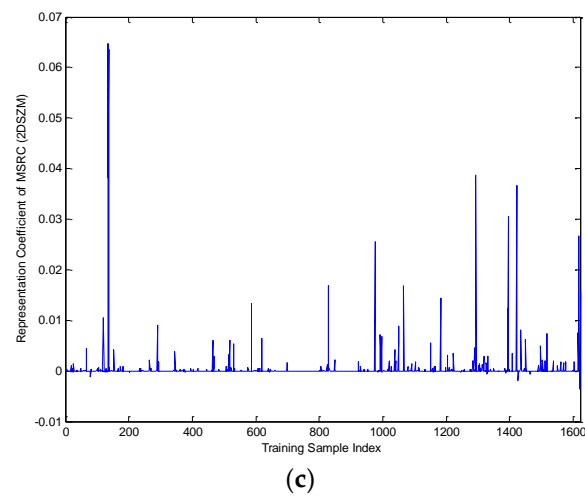
**Figure 6.** Test sample BMP2-9563 ( $172.50^\circ$ ).



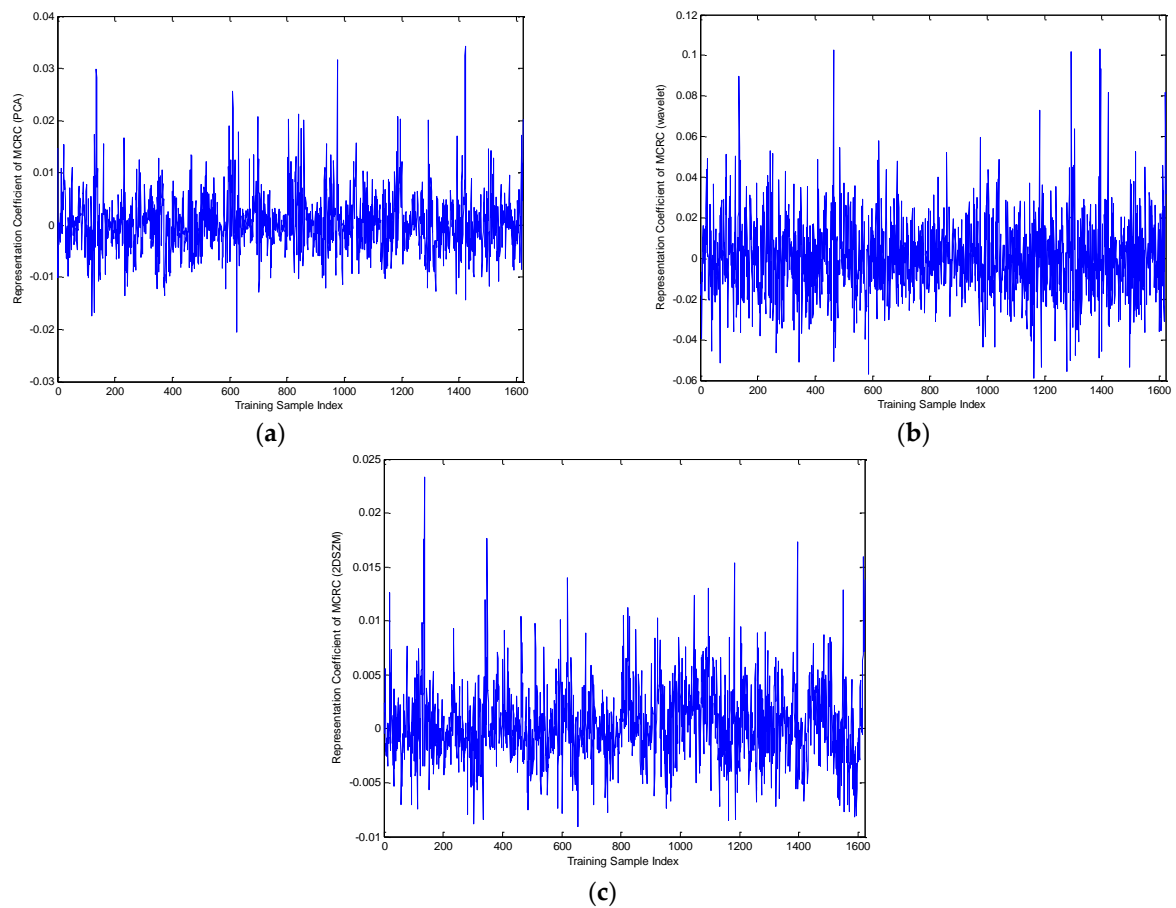
**Figure 7.** (a) The second-stage representation coefficient of principal component analysis (PCA) features of TSMRL; (b) the second-stage representation coefficient of wavelet features of TSMRL; (c) the second stage representation coefficient of 2DSZM features of TSMRL.



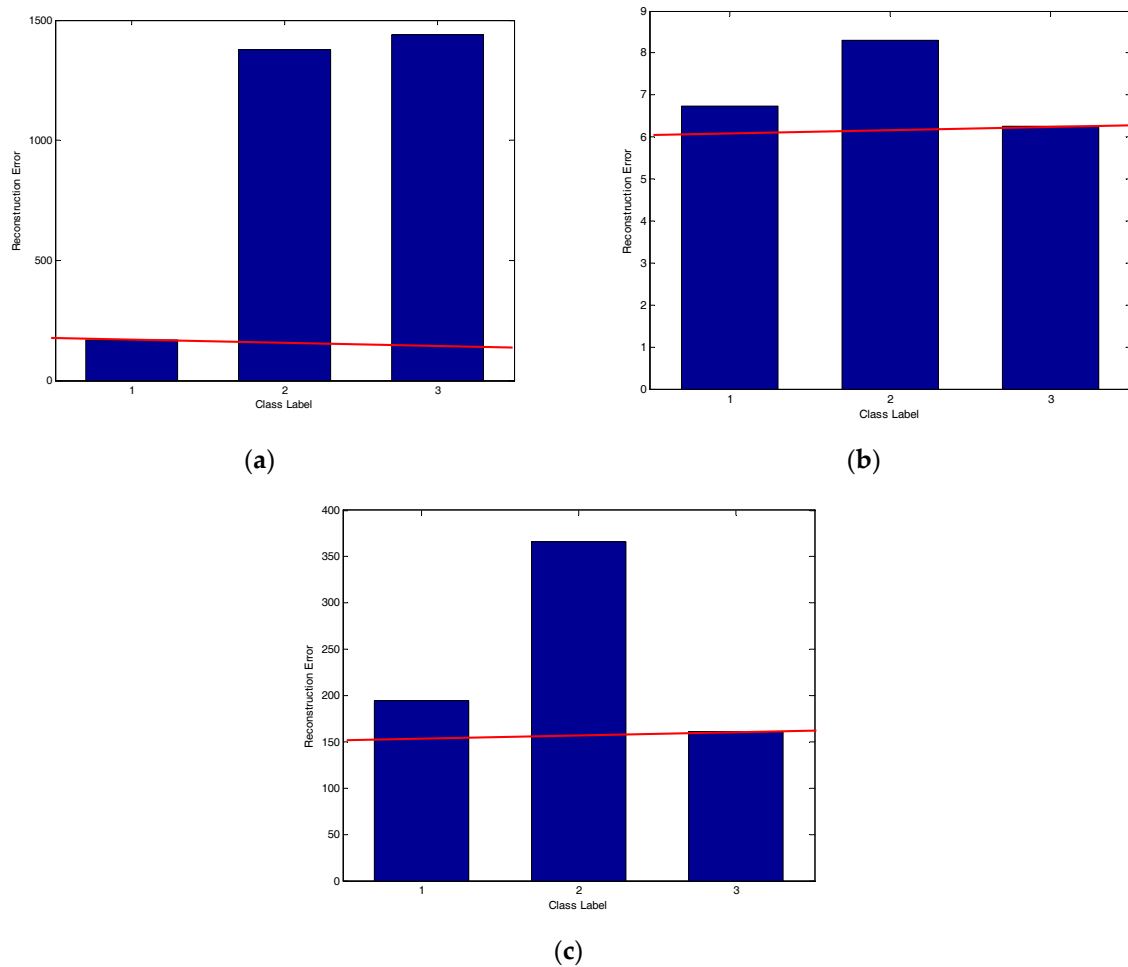
**Figure 8.** Cont.



**Figure 8.** (a) Representation coefficient of PCA features of multi-task sparse representation classifier (MSRC); (b) representation coefficient of wavelet features of MSRC; (c) representation coefficient of 2DSZM features of MSRC.



**Figure 9.** (a) Representation coefficient of PCA features of multi-task collaborative representation classifier (MCRC); (b) representation coefficient of wavelet features of MCRC; (c) representation coefficient of 2DSZM features of MCRC.



**Figure 10.** (a) Reconstruction error of TSMRL; (b) reconstruction error of MSRC; (c) reconstruction error of MCRC.

For MSRC and MCRC, coding was performed across the global azimuth range of 0–360°. For TSMRL, after the first stage  $\ell_{2,1}$ -norm regularized multi-task joint sparse learning, a subset of training samples was selected. The training samples in the selected subset became the effective atoms, while the other atoms were set to zero vectors. In this way, we constructed a new dictionary, which the second stage was based upon. These selected neighbor atoms play an important role for correct classification. In summary, the experimental results shown in Figures 7–10 demonstrate that TSMRL can prevent the interference of atoms far away from the test sample, and improve recognition performance.

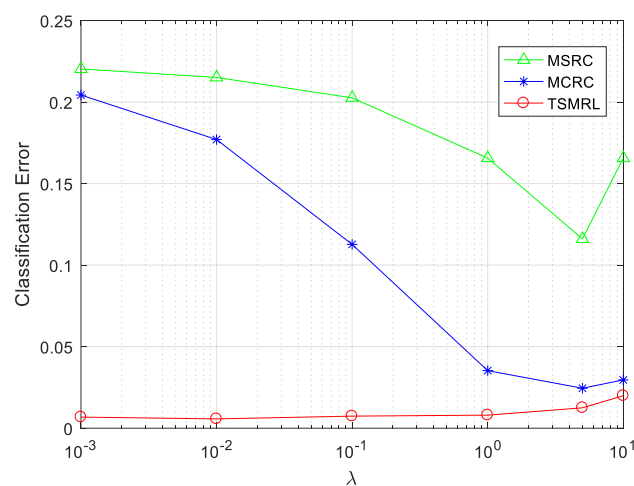
#### 4.2.3. Classification Performance with Regularization Parameter Value Change

There are two regularization parameters,  $\lambda$  and  $\rho$ , in TSMRL. The values of the two regularization parameters have an important impact on classification performance.

There is only one balance parameter in both MSRC and MCRC. In order to compare TSMRL with MSRC and MCRC, two regularization parameters in TSMRL take the same value as those in the other two algorithms. In this experiment, the value of  $Q$  was fixed to 10, and the regularization parameter values were set to  $10^{-3}$ ,  $10^{-2}$ , 0.1, 1, 5 and 10 respectively. Figure 11 and Tables 3 and 4 show the classification error variation diagram, the error rate table, and the confusion matrix, respectively. From the curve of Figure 11, it can be seen that MSRC and MCRC are sensitive to changes of regularization parameter values, and the recognition error rates tend to descend with the increase of the value from  $10^{-3}$  to 5. However, TSMRL is rather smooth with changes of regularization parameter



values, and the error rate does not change obviously with the variety of parameter values, which shows that TSMRL had better robustness to various regularization parameters. This is very important in practical applications because you do not always know the ideal regularization parameters. Taking the regularization parameter value 0.1 as an example, the classification error rates corresponding to MSRC, MCRC and TSMRL were 20.26%, 11.27% and 0.74%, respectively. The classification performance of TSMRL was much better than that of MSRC and MCRC. The entire statistical classification error rates are shown in Table 3. It is clear that TSMRL has the highest classification accuracy and small fluctuation. By contrast, the classification performance of MSRC and MCRC is not stable. In addition, it can be seen that when the regularization parameter value is  $10^{-2}$ , TSMRL achieves the minimum error rate of 0.57%, i.e., the correct recognition rate is as high as 99.43%. As shown in Table 4, with the regularization parameter value  $10^{-2}$  as an example, the number of correct classified three targets obtained by TSMRL is better than that of MSRC and MCRC under the same conditions. The number of BMP2 correctly identified by TSMRL is 583, which is significantly better than the 453 of MSRC and 560 of MCRC. The results of these experiments illustrate that TSMRL had the best discrimination ability among the three algorithms.



**Figure 11.** Classification performance of the three algorithms with different values of regularization parameter  $\lambda$ .

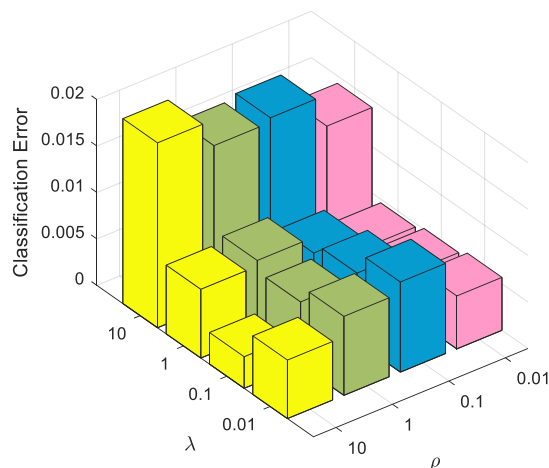
**Table 3.** Classification error rate of the three algorithms with regularization parameter  $\lambda$ .

$\lambda$	Method		
	TSMRL	MSRC	MCRC
$10^{-3}$	0.68%	22.03%	20.43%
$10^{-2}$	0.57%	21.51%	17.70%
0.1	0.74%	20.26%	11.27%
1	0.80%	16.56%	3.53%
5	1.25%	11.61%	2.45%
10	1.99%	16.56%	2.96%

**Table 4.** Confusion matrix of the three algorithms under different regularization parameter values.

Method	Ground Truth	$\lambda = 10^{-3}$			$\lambda = 10^{-2}$		
		BMP2	BTR70	T72	BMP2	BTR70	T72
TSMRL	BMP2	585	0	2	586	0	1
	BTR70	5	579	4	4	581	3
	T72	1	1	580	1	1	580
MSRC	BMP2	433	91	63	439	86	62
	BTR70	66	487	35	64	495	29
	T72	56	76	450	57	80	445
MCRC	BMP2	548	0	39	552	0	35
	BTR70	168	274	146	145	317	126
	T72	6	0	576	5	0	577
Method	Ground Truth	$\lambda = 0.1$			$\lambda = 1$		
		BMP2	BTR70	T72	BMP2	BTR70	T72
TSMRL	BMP2	583	0	4	583	1	3
	BTR70	6	580	2	8	578	2
	T72	1	0	581	0	0	582
MSRC	BMP2	453	73	61	484	64	39
	BTR70	53	510	25	54	510	24
	T72	65	79	438	55	55	472
MCRC	BMP2	560	0	27	562	2	23
	BTR70	96	421	71	15	552	21
	T72	4	0	578	1	0	581
Method	Ground Truth	$\lambda = 5$			$\lambda = 10$		
		BMP2	BTR70	T72	BMP2	BTR70	T72
TSMRL	BMP2	581	1	5	579	1	7
	BTR70	6	574	8	17	563	16
	T72	1	1	580	2	0	580
MSRC	BMP2	536	32	19	539	9	39
	BTR70	53	505	30	122	381	85
	T72	28	42	512	14	22	546
MCRC	BMP2	554	7	26	542	11	34
	BTR70	0	580	8	0	585	3
	T72	0	2	580	1	3	578

TSMRL uses two-stage multi-task representation learning. The first stage is the  $\ell_{2,1}$ -norm regularized multi-task sparse learning, the balance parameter of which is  $\lambda$ . The second stage is the Frobenius norm regularized multi-task collaborative learning, and the balance parameter is  $\rho$ . In previous experiments,  $\lambda$  and  $\rho$  took the same value. However, in practice their values are often inconsistent. Therefore, it is necessary to analyze the classification performance when both values are not equal. We set  $\lambda$  and  $\rho$  for different values and carried out the classification experiment using TSMRL. The results are shown in Figure 12. As can be seen from Figure 12, the error rates of TSMRL were lower than 2% when  $\lambda$  and  $\rho$  are in the range of  $10^{-2}$ –10. The results further illustrate TSMRL's robustness to a variety of regularization parameter values. On the other hand, when  $\lambda$  is 0.1 and  $\rho$  is 10, TSMRL achieves the minimum error rate of 0.34%, i.e., the correct recognition rate was as high as 99.66%. These experimental results support the advantages of the proposed algorithm.

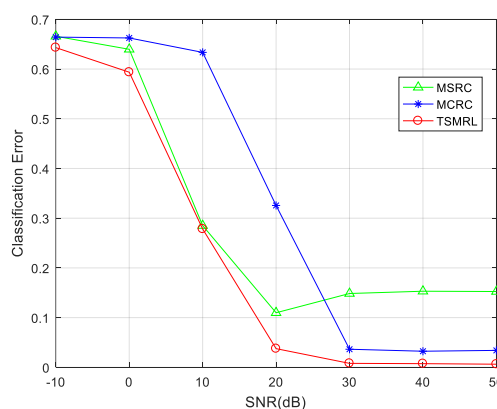


**Figure 12.** Classification errors of TSMRL with different regularization parameters.

Furthermore, we compared the performance of the proposed method with recent SAR target images classification methods. Under the same dataset, the method in [19] achieved a recognition rate of 96.48%; the approach in [32] achieved a recognition rate of 98.63%. The recognition rate of the proposed method reached 99.66%, which was better than the above methods.

#### 4.2.4. Robustness to Noise

Due to the complex imaging environment of the battlefield, it is important to investigate the anti-noise performance of the proposed algorithm. Gaussian noises of different levels with varying signal-to-noise ratios (SNR) were added into testing samples, thereby enabling evaluation of the robustness of the algorithm. Figure 13 shows the classification performance curves of the three algorithms at different SNRs. From the figure, we can see that the error rates of all algorithms decrease with higher SNR. Nevertheless, the TSMRL algorithm has the minimal error rate among all methods for each SNR level. Furthermore, an error rate of 3.67% can be achieved under 20 dB SNR by the TSMRL algorithm, and an error rate of 0.74% can be obtained under 40 dB SNR level, which illustrates its robustness to noise corruption.



**Figure 13.** Performance of three algorithms with varying signal-to-noise ratios (SNRs).

#### 4.2.5. Experiments Conducted with Depression Angle Variations

In practical applications, SAR images of a target are not collected at a fixed depression angle. When the depression angle of the target changes, the SAR image varies as well. In other words, a change in depression angle also affects the classification performance of the algorithm. Thus, it is necessary to

analyze the classification performance with depression angle variations. In our experiments, the three algorithms were evaluated with a large depression angle variation. Three targets, BRDM2, 2S1, and ZSU23/4 were employed, as shown in Figure 14. SAR images of these targets are shown in Figure 15. The dataset used in the experiment is given in Table 5. SAR images collected at a  $17^\circ$  depression angle were used for training, while images collected at  $30^\circ$  and  $45^\circ$  were employed for testing. The experimental results are shown in Table 6.

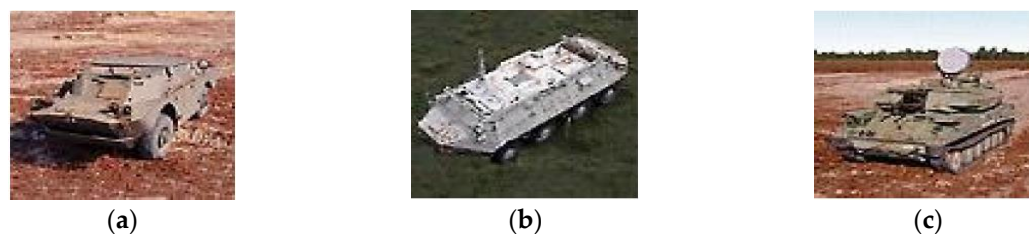


Figure 14. Optical images of three targets. (a) BRDM2; (b) 2S1; (c) ZSU23/4.

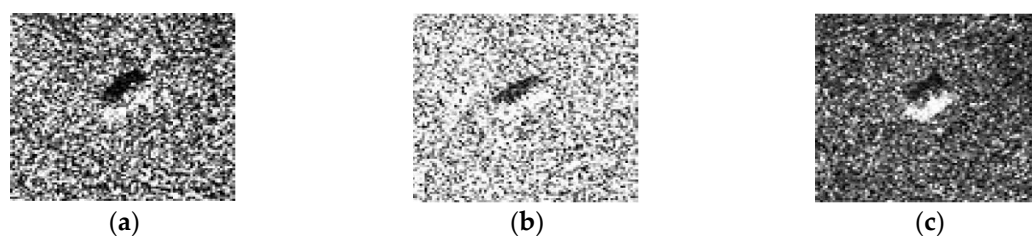


Figure 15. SAR images of three targets. (a) BRDM2; (b) 2S1; (c) ZSU23/4.

Table 5. Dataset used in experiment.

	BRDM2	2S1	ZSU23/4
Training Set ( $17^\circ$ )	298	299	299
Testing Set ( $30^\circ$ )	287	288	288
Testing Set ( $45^\circ$ )	303	303	303

Table 6. Classification error under different depression angles of the three methods.

Depression	Method		
	TSMRL	MSRC	MCRC
$30^\circ$	5.33%	28.85%	8.11%
$45^\circ$	44.11%	63.81%	59.52%

As shown in Table 6, with a  $30^\circ$  depression angle, the error rate of TSMRL was 5.33%, obviously lower than those of MSRC and MCRC. With the depression angle increasing to  $45^\circ$ , the recognition performance of the three methods all declined. This was mainly because SAR images of the targets changed severely with an angle gap of  $28^\circ$  between training and test samples. However, TSMRL could still achieve the minimum classification errors, which were 19.70% and 15.41% lower than those of MSRC and MCRC, respectively.

The experimental results verify that TSMRL is most robust when it comes to the depression variation among the three algorithms. This is because of the two-stage strategy reducing confusion caused by large depression variation, which ensures accurate multi-task learning.

## 5. Conclusions

In this paper, we proposed a two-stage multi-task representation learning model to classify SAR target images. The proposed method could not only make use of multi-feature collaborative discriminative ability for targets, but could also integrate the representation learning capacity of sparse representation and collaborative representation. The first stage  $\ell_{2,1}$ -norm regularized multi-task sparse learning selects the nearest subset of the test sample from the perspective of representation learning, which greatly reduces the interference caused by outliers. In this way, a new dictionary is obtained. Through the second stage of multi-task collaborative representation learning, the rational representation weight coefficients are assigned to each atom in the new dictionary. Subsequently, the label of the test sample can be inferred accurately based on the criterion of the minimum reconstruction error. Moreover, because the second-stage multi-task collaborative representation has a closed form solution, the computational complexity of the whole algorithm is mainly embodied in the first stage of multi-task sparse representation learning. The experimental results show that the classification performance of the proposed algorithm is much better than that of multi-task sparse representation learning and multi-task collaborative representation learning, and the algorithm is robust for a wide range of regularization parameters.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China under Grants No. 61301224, No. 61501068, and No. 41474023. This work was also supported by the Basic and Advanced Research Project in Chongqing under Grants No. cstc2017jcyjA1378 and No. cstc2016jcyjA0134.

**Author Contributions:** X.Z. conceived and designed the experiments, analyzed the data, and wrote part of the paper. Y.W. and Z.T. performed the experiments and wrote part of the paper. D.L., S.L., T.W. and Y.L. gave some revision suggestions for the paper.

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

## References

1. Wang, H.; Chen, Z.; Zheng, S. Preliminary research of low-RCS moving target detection based on Ka-band video SAR. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 811–815. [[CrossRef](#)]
2. Yu, Q.; Clausi, D.A. SAR sea-ice analysis based on iterative region growing using semantics. *IEEE Trans. Geosci. Remote Sens.* **2007**, *12*, 3919–3931. [[CrossRef](#)]
3. Velotto, D.; Bentes, C.; Tings, B.; Lehner, S. First comparison of Sentine-1 and TerraSAR-X data in the framework of maritime targets detection: South Italy Case. *IEEE J. Ocean. Eng.* **2016**, *4*, 993–1006. [[CrossRef](#)]
4. El-Darymli, K.; Gill, E.W.; McGuire, P.; Power, D.; Moloney, C. Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review. *IEEE Access* **2016**, *4*, 6014–6058. [[CrossRef](#)]
5. Novak, L.M.; Owirka, G.J.; Netishen, C.M. Performance of a high-resolution polarimetric SAR automatic target recognition system. *Lincoln Lab. J.* **1993**, *6*, 11–23.
6. Ma, C.; Wen, G.J.; Ding, B.Y.; Zhong, J.R.; Yang, X. Three-dimensional electromagnetic model-based scattering center matching method for synthetic aperture radar automatic target recognition by combining spatial and attributed information. *J. Appl. Remote Sens.* **2016**, *10*, 122–134. [[CrossRef](#)]
7. Wagner, S.A. SAR ATR by a combination of convolutional neural network and support vector machines. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 2861–2872. [[CrossRef](#)]
8. Peng, C.; Kang, Z.; Cheng, Q. Integrating feature and graph learning with low-rank representation. *Neurocomputing* **2017**, *249*, 106–116. [[CrossRef](#)]
9. Zhang, Z.; Xu, Y.; Yang, J.; Li, X.; Zhang, D. A survey of sparse representation: Algorithms and applications. *IEEE Access* **2017**, *3*, 490–530. [[CrossRef](#)]
10. Bao, B.K.; Zhu, G.; Shen, J.; Yan, S. Robust image analysis with sparse representation on quantized visual features. *IEEE Trans. Image Process.* **2013**, *22*, 860–871. [[CrossRef](#)] [[PubMed](#)]
11. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [[CrossRef](#)] [[PubMed](#)]
12. Shahnawazuddin, S.; Sinha, R. Sparse coding over redundant dictionaries for fast adaptation of speech recognition system. *Comput. Speech Lang.* **2016**, *43*, 1–17. [[CrossRef](#)]

13. Fang, L.; Wang, C.; Li, S.; Benediktsson, J. Hyperspectral image classification via multiple-feature-based adaptive sparse representation. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1646–1657. [[CrossRef](#)]
14. Zhang, H.; Nasrabadi, N.M.; Zhang, Y.; Huang, T.S. Multi-view automatic target recognition using joint sparse representation. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 2481–2497. [[CrossRef](#)]
15. Cheng, J.; Li, L.; Li, H.; Wang, F. SAR target recognition based on improved joint sparse representation. *EURASIP J. Adv. Signal Process.* **2014**, *2014*, 87. [[CrossRef](#)]
16. Dong, G.; Kuang, G.; Wang, N.; Zhao, L.; Lu, J. SAR target recognition via joint sparse representation of monogenic signal. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2015**, *8*, 3316–3328. [[CrossRef](#)]
17. Dong, G.; Kuang, G. SAR target recognition via sparse representation of monogenic signal on grassmann manifolds. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 1308–1319. [[CrossRef](#)]
18. Liu, M.; Wu, Y.; Zhao, W.; Zhang, Q.; Li, M.; Liao, G. Dempster-Shafer fusion of multiple sparse representation and statistical property for SAR target configuration recognition. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1106–1110. [[CrossRef](#)]
19. Sun, Y.; Du, L.; Wang, Y.; Wang, Y.; Hu, J. SAR automatic target recognition based on dictionary learning and joint dynamic sparse representation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1777–1781. [[CrossRef](#)]
20. Song, S.; 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](#)]
21. Liu, H.; Jiu, B.; Li, F.; Wang, Y. Attributed scattering center extraction algorithm based on sparse representation with dictionary refinement. *IEEE Trans. Antennas Propag.* **2017**, *65*, 2604–2614. [[CrossRef](#)]
22. Chi, Y.; Porikli, F. Classification and boosting with multiple collaborative representations. *IEEE Comput. Sci.* **2014**, *36*, 1519–1531. [[CrossRef](#)] [[PubMed](#)]
23. Li, W.; Du, Q.; Zhang, F.; Hu, W. Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2014**, *12*, 389–393. [[CrossRef](#)]
24. Leiva-Murillo, J.M.; Gomez-Chova, L.; Camps-Valls, G. Multitask remote sensing data classification. *IEEE Geosci. Remote Sens. Lett.* **2013**, *51*, 151–161. [[CrossRef](#)]
25. Yuan, X.T.; Yan, S. Visual classification with multitask joint sparse representation. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3493–3500.
26. Fang, L.; Li, S. Face recognition by exploiting local Gabor features with multitask adaptive sparse representation. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2605–2615. [[CrossRef](#)]
27. Li, J.; Zhang, H.; Zhang, L.; Huang, X.; Zhang, L. Joint collaborative representation with multitask learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5923–5936. [[CrossRef](#)]
28. Luo, Y.; Tao, D.C.; Geng, B.; Xu, C.; Maybank, S.J. Manifold regularized multitask learning for semi-supervised multilabel image classification. *IEEE Trans. Image Process.* **2013**, *22*, 523–537. [[CrossRef](#)] [[PubMed](#)]
29. Luo, Y.; Wen, Y.G.; Tao, D.C.; Gui, J.; Xu, C. Large margin multi-modal multi-task feature extraction for image classification. *IEEE Trans. Image Process.* **2016**, *25*, 414–428. [[CrossRef](#)]
30. Ding, J.; Chen, B.; Liu, H.; Huang, M. Convolutional neural network with data augmentation for SAR target recognition. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 364–368. [[CrossRef](#)]
31. Chen, S.; Wang, H.; Xu, F.; Jin, Y.Q. Target classification using the deep convolutional networks for SAR images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4806–4817. [[CrossRef](#)]
32. Zhang, X.; Liu, Z.; Liu, S.; Li, D.; Jia, Y.; Huang, P. Sparse coding of 2D-slice Zernike moments for SAR ATR. *Int. J. Remote Sens.* **2017**, *38*, 412–431. [[CrossRef](#)]
33. Wu, X.; Li, Q.; Xu, L.; Chen, K.; Yao, L. Multi-feature kernel discriminant dictionary learning for face recognition. *Pattern Recognit.* **2017**, *66*, 404–411. [[CrossRef](#)]
34. Donoho, D.L.; Elad, M. Optimal sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2197–2202. [[CrossRef](#)] [[PubMed](#)]
35. Candes, E.J.; Tao, T. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory* **2006**, *52*, 5406–5425. [[CrossRef](#)]
36. Zhao, P.; Yu, B. On model selection consistency of Lasso. *J. Mach. Learn. Res.* **2006**, *7*, 2541–2563.
37. Zhang, L.; Yang, M.; Feng, X. Sparse representation or collaborative representation: Which helps face recognition? In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 471–478.



38. Liu, Z.Y.; Chiu, K.C.; Xu, L. Improved system for object detection and star/galaxy classification via local subspace analysis. *Neural Netw.* **2003**, *16*, 437–451. [[CrossRef](#)]
39. Vural, V.; Fung, G.; Krishnapuram, B.; Dy, J.G.; Rao, B. Using local dependencies within batches to improve large margin classifiers. *J. Mach. Learn. Res.* **2009**, *10*, 183–206.
40. Xu, Y.; Zhang, D.; Yang, J.; Yang, J.Y. A two-phase test sample sparse representation method for use with face recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 1255–1262. [[CrossRef](#)]
41. Argyriou, A.; Evgeniou, T.; Pontil, M. Multitask Feature Learning. In Proceedings of the 19th International Conference on Neural Information Processing Systems (NIPS'06), Vancouver, BC, Canada, 4–7 December 2006; pp. 41–48.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).