EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

Selecting pseudo supervision for unsupervised domain adaptive SAR target classification

Lingjun Zhao[1*], Qishan He[1], Ding Ding[2], Siqian Zhang[1], Gangyao Kuang[1] and Li Liu[3]

*Correspondence: nudtzlj@163.com

[1] State Key Laboratory of Complex Electromagnetic Environment Effects On Electronics and Information System, College of Electronic Science, National University of Defense Technology, Changsha 410073, China
[2] Center for Teaching and Research Service, National University of Defense Technology, Changsha 410073, China
[3] College of System Engineering, National University of Defense Technology, Changsha 410073, China

Abstract

In recent years, deep learning has brought significant progress for the problem of synthetic aperture radar (SAR) target classification. However, SAR image characteristics are highly sensitive to the change of imaging conditions. The inconsistency of imaging parameters (especially the depression angle) leads to the distribution shift between the training and test data and severely deteriorates the classification performance. To address this problem, in this paper we propose an unsupervised domain adaptation method based on selective pseudo-labelling for SAR target classification. Our method directly trains a deep model using the data from the target domain by generating pseudo-labels in the target domain. The key idea is to iteratively select valuable samples from the target domain and optimize the classifier. In each iteration, the breaking ties (BT) criterion is adopted to select the best samples with the highest scores of relative confidence. Besides, to avoid error accumulation in the iterative process, class confusion regularization is used to improve the accuracy of pseudo-labelling. Our method is compared with state-of-the-art methods, including supervised classification and unsupervised domain adaptation methods, over the moving and stationary target acquisition and recognition (MSTAR) dataset. The experimental results demonstrate that the proposed method can achieve better classification performance, especially when the difference of depression angles of the source and target domain images is large. Besides, our method also shows its superiority under limited-sample conditions.

Keywords: SAR image, Target classification, Unsupervised domain adaptation, Selective pseudo-labelling

1 Introduction

Synthetic aperture radar (SAR) has been widely used in military and civilian applications since it can provide high-resolution imagery in all-weather and day-night time. SAR target recognition is a typical image pattern recognition problem, which aims to classify the target images into different classes or types. Traditionally, a target classification algorithm generally consists of three steps: preprocessing, extracting features, and classifying. Deep learning (DL) has been widely used in the computer vision field in recent years. The automatic feature-extraction ability has attracted much attention in SAR automatic target recognition (ATR). Chen et al. [1] used a convolution layer instead

of a full connection layer to realize feature classification in a convolutional neural network (CNN), and proposed a full convolution neural network called A-Convnet, which achieved a classification accuracy of 99.13% under the standard operating conditions of MSTAR data. Wagner et al. [2] used a support-vector machine to classify features extracted by CNN and proposed the structure of CNN-SVM (support-vector machine). At present, most of the DL-based methods can achieve an accuracy of more than 99% under the standard operating conditions of the MSTAR dataset. Target classification using limited data is also a research hotspot in SAR ATR. Many techniques based on adjusting model structure [3], transfer learning [4, 5], data augmentation [6] are adopted to mitigate the overfitting problem caused by the limited training data.

With the development of machine learning in the field of SAR target recognition, the existing methods can achieve better target recognition performance under sufficient and insufficient training samples. However, most methods usually only use the labelled data in the source domain to train a classifier in a supervised way, so it is assumed that the training data and test data come from the same or similar distribution. When this assumption fails, the model learned on the training set is difficult to achieve good performance on the test set, which is caused by inconsistent distribution. SAR image is highly sensitive to imaging conditions. The inconsistency of imaging parameters such as depression angle, azimuth angle, and radar band will lead to the distribution mismatch between the training and test samples [7]. As shown in Fig. 1, depression angle differences cause visual discrepancy of a certain vehicle target. In this case, the classifier trained by the source domain is difficult to obtain accurate classification results on the test set. Therefore, how to overcome the degradation of model generalization performance caused by different imaging conditions is an urgent problem to be solved.

Deep unsupervised domain adaptation techniques can be employed in addressing such domain shift problems. Existing deep unsupervised domain adaptation approaches mainly learn domain-invariant features from the labelled source domain data and the unlabelled target domain data via explicit source and target data distributions [8–15]. Although impressive performance has been achieved in prior works, we argue that such methods designed for optical image classification tasks cannot achieve good results in SAR ATR tasks for two reasons. First, due to the high cost and difficulty of manual annotation of SAR data, deep neural networks turn out to be overfitting to a certain extent with limited labelled training data. Second, alignment between the source and target domains in an unsupervised manner is not efficient enough to handle large imaging discrepancy of SAR images. Furthermore, it has been justified that domain shift can be

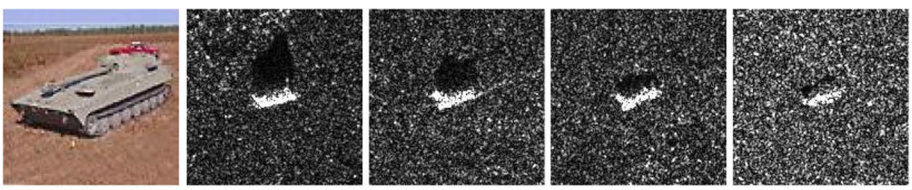

Fig. 1 SAR images of the 2S1 vehicle target at different depression angles. **a** Optical image. **b** SAR image with 15° depression angle. **c** SAR image with 17° depression angle. **d** SAR image with 30° depression angle. **e** SAR image with 45° depression angle

unified through a classifier trained on both the source and target domain data in a high-dimensional feature space due to the dimensionality blessing [16]. Hence, it is beneficial to considering the labelled target domain as an auxiliary training data.

To improve the domain adaptation ability of SAR image targets in different imaging conditions, an unsupervised domain adaptation approach based on selective pseudo-labelling is proposed in this paper. The idea of pseudo-labelling is to first train the initial classification model in the source domain, then test the samples in the target domain to generate pseudo-labels. In each iteration, the BT criterion is adopted to select the best samples with the highest scores of relative confidences. Next, target samples with pseudo-labels are added to the training set for supervised training. This process is repeated until the generated pseudo-labels do not change. On one hand, the pseudo-labelled samples in the target domain can expand the training set and inhibit the overfitting of the model. On the other hand, it can directly adapt the model to the potential feature distribution of the target domain. However, the pseudo-labelling strategy is extremely dependent on the initial pseudo-labels. If the initial pseudo-labels are wrongly assigned, it is easy to result in error accumulation and cause the model fall into local optimal solution. To avoid the error accumulation in the iterative process, class confusion regularization is used to improve the accuracy of pseudo-labelling. We conducted the experiments on SAR images with different depression angles to explore the cross-domain classification capability of our method. Based on the MSTAR dataset, six configurations of depression angles are considered to construct the source domain and target domain data for target classification. Our method is compared with state-of-the-art methods, including supervised classification and unsupervised domain adaptation methods. Classification results and feature visualization based on t-distributed stochastic neighbor embedding (t-SNE) [17] demonstrate that the proposed method can achieve better classification performance, especially when the difference of depression angles of the source and target domain images is large. Besides, our method also shows its superiority under few-sample conditions.

2 Unsupervised deep domain adaptation

In computer vision, a common assumption behind image classification task is that the source domain and the target domain data have similar or the same support [18]. However, in many real-world scenarios, this assumption fails since there may be no overlapping features across the source domain and target domain. We give formal related notations mathematically to define the problem. In supervised learning, given a source training sample set $X^s = \{(x_i^s)\}_{i=1}^N$ and corresponding label set $Y^s = \{(y_i^s)\}_{i=1}^N$, the goal of a learner usually consists of finding a good hypothesis function h that captures in the best way possible the relation between X and Y. This relationship often extends beyond the training instances to test instances $X^t = \{(x_i^t)\}_{i=1}^M$ drawn from the same probability. However, when such probability identity does not hold, a classifier trained on the labelled source domain suffers from significant performance drop when directly applied to the target domain, as shown in Fig. 2. To be specific, the marginal distributions of source and target domains are different, i.e., $p(x^s) \neq p(x^t)$, but the conditional probability distributions are identical, i.e., $p(y^s|x^s) = p(y^t|x^t)$.

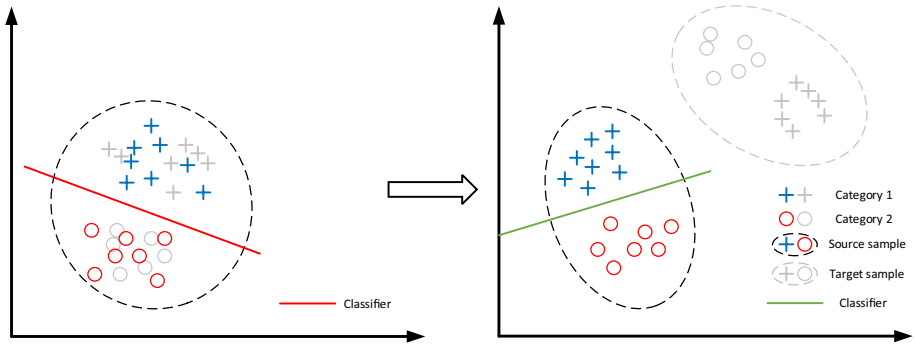

Fig. 2 Illustration of classifier and samples under two cases. **a** Distributions of the source and target domains keep the same. **b** Distributions of the source and target domain is quite different

From this perspective, unsupervised deep domain adaptation approaches consider learning domain-invariant features through labelled source data and unlabelled target data in an end-to-end framework. DDC (deep domain confusion) [8] aims to learn transferable features by matching kernel embedding in reproducing kernel Hilbert space (RKHS) calculated from two distributions. DAN (deep adaptation networks) [9] improves the DDC domain metric by replacing it with a multi-kernel variant. Deep Coral [10] utilizes the idea of Coral to learn features with the same second order statistical property. JAN (joint adaptation network) [11] aligns features activated from multilayers using joint maximum mean discrepancy. These methods are based on different advanced domain discrimination metrics. Meanwhile, adversarial training strategy of generative adversarial networks [19] are employed in domain adaptation to learn domain-invariant features. The key is to play a 'minmax' game over the features from multilayers output activation. DaNN (domain adversarial neural network) [20] proposes the gradient reversal layer, through which features from two domains are made as indistinguishable as possible during the gradient backpropagation. The success of gradient reversal layer lies on that the 'minmax' process of the feature extraction network can be conducted in once backpropagation. CDAN (conditional domain adaptation network) [12] improves the adversarial training by capturing the mutual covariance between features and classifier's output. More recently, MCD (maximum classifier discrepancy) [13] attempts to align distributions of the source and target domains by utilizing the task-specific decision boundaries. Two classifiers are trained to maximize the discrepancy to detect target samples that are far from the support of the source data, and a feature generator learns to generate target features near the support to minimize the discrepancy.

It is noticed that unsupervised domain adaptation assumes the access of labelled data only from the source domain and unlabelled data from the target domain, hence it follows transductive learning paradigm. One effective method is to train networks with samples from both domains in a supervised manner. Pseudo-labelling has been used in many approaches to help address the lack of labelled data, such as semi-supervised learning [21] and few-shot learning [16, 22]. Two categories, i.e., hard labelling [23, 24] and soft labelling [25] have been employed in many existing works. The main idea of hard pseudo-labelling is to assign a pseudo-label to each unlabelled instance and then train classifiers with an augmented labelled training set. It is supposed that the parameters of networks trained based on hard labelling tends to be stuck in local maximum

since it does not consider each target sample's confidence. The strategy of soft labelling assigns the conditional probability of each class given a target sample [25]. To address the mis-labelling issue, selective pseudo-labelling is another effective method, which selects part of the unlabelled samples in some sort of order to assign pseudo-labels. One key factor is how to make criterion of sample selection for pseudo-labelling. An easy-to-hard strategy is employed in [26]. Target samples whose similarity scores are higher than a certain threshold are selected for pseudo-labelling and this threshold is updated after each iteration of learning so that more unlabelled target samples can be selected.

Although impressive performance of unsupervised deep domain adaptation has been achieved in optical image classification, they are rarely used for SAR target recognition. In this paper, we propose a novel domain adaptation approach based on selective pseudo-labelling aimed to address the imaging discrepancy for SAR target classification. The main contributions of this article are as follows.

(1) SAR images are highly susceptible to imaging conditions, which causes recognition degradation on deep learning models. We firstly investigate the model's generalization ability across images captured from different depression angles, which is ignored in previous studies on SAR-ATR.
(2) A selective pseudo-labelling strategy is introduced into the domain adaptation method. This strategy not only implicitly conducts feature alignment without moment distribution or adversarial learning, but also boosts the model's generalization under limited training data.
(3) To avoid error accumulation of pseudo-labelling, class confusion loss is introduced into the iterative process as a regularization term, to enhance the pseudo-labelling accuracy gradually in each iteration.
(4) Our proposed method obtains an obvious improvement over the compared domain adaptation methods across different depression angle data in MSTAR datasets. Furthermore, our method is more suitable in SAR-ATR under training conditions of limited samples.

3 Method

Classical domain adaptation methods usually learn domain-invariant features by directly or indirectly aligning the distributions of the source domain and target domain. However, for high-dimensional classification neural networks, it is possible to learn a classifier with certain generalization performance when labels in the source domain and target domain are available. Since unsupervised domain adaptation has assumed that there are no available labels for the target domain, we propose a SAR image domain adaptation approach based on selective pseudo-labelling with class confusion regularization (SPL-CCR).

3.1 Overall architecture of SPL-CCR

The diagram of SPL-CCR is shown in Fig. 3. SPL-CCR organize two main steps in an iterative learning strategy to generate pseudo-labels from the target domain. In the first step, we train the classification network with the labelled source data and classify

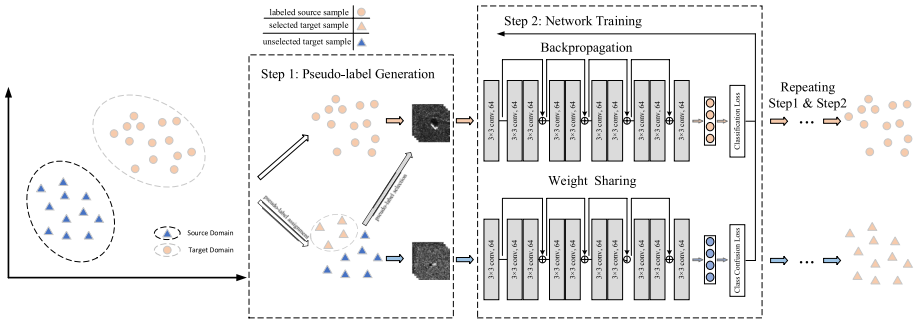

Fig. 3 The diagram of SPL-CCR for SAR target classification

Table 1 Domain adaptation model based on pseudo-labelling

Variable	Algorithm
X_s: Dataset in the source domain X_t: Dataset in the target domain X_l: Labelled training data set initialized as the source data X_s G: Classifier trained on T Q: Selective pseudo-labelling strategy	*Initialization* Train the classifier G on the labelled training data set X_l Assign pseudo-labels for all the data in X_t using the classifier G Repeat: Select pseudo-labelled samples in X_t according to Q Add samples selected in (3) to X_l Retrain the classifier G Until the iteration condition is met or the pseudo-labels no longer change

the unlabelled target samples. Then, a fraction of target samples is assigned with pseudo-labels according to selection strategy. In the second step, the classification network is trained with two input streams. One of is composed of the source domain samples and the selected target domain samples. The other is composed of the unselected unlabelled target domain samples. The classifier gets stronger after learning from the former trainset in a supervised way. The second train set is also sent to the network to calculate the class confusion loss, which plays the role of regularization to alleviate the mis-labelling issue. Specially, we adopt ResNet18 [27] and a custom linear layer as our feature extraction network and classifier.

The key idea of pseudo-labelling is to iteratively establish valuable sample set from the target domain and optimize the classifier. Different from traditional supervised learning and domain adaptation, the classifier will be trained by data from both the source domain and target domain. To better describe the learning process, we introduce the model (T, X_s, X_t, G, Q). G is a supervised classification network, which is trained using the training set T. X_s is the labelled source domain dataset. X_t is the unlabelled target domain dataset used to provide pseudo-labels. Q refers to the pseudo-labelling strategy, which is used to select valuable samples and automatically assign pseudo-labels.

As shown in Table 1, pseudo-labelling-based domain adaptation is trained in an iterative way. In the beginning, the training set is initialized as the source domain data X_s, which have been used to train the classifier G. Next, in each iteration, the most valuable samples from the unlabelled target domain are selected and each is assigned a pseudo-label. The selected samples with pseudo-labels are added to the training set X_l and the

updated training set is used to retrain the classifier G. The process is repeated until the iteration condition is reached or the pseudo-labels no longer change.

3.2 Pseudo-labelling selection strategy

Wang et al. [28] proved that it is beneficial to select some valuable pseudo-labelled samples from the target domain as part of the training set instead of using all the pseudo-labelled samples. When domain discrepancy exists, the classifier trained on the initial training set, i.e., the labelled source domain data, usually shows low accuracy on the target domain. Therefore, it is better to use only a small fraction of the target samples at the beginning.

It is supposed to select the samples with high probability to be correctly classified by the current network to alleviate the mis-labelling issue. We adopt the BT criterion [29] which is inspired to the multiclass-level uncertainty for classification with SVMs. The main idea of the BT criterion is that the best sample has the least uncertainty between the two classes to which it is most likely to belong. The relative confidence of the target sample given by the current network is defined as follows

$$U_i = \max_{c\in\Omega} p(\widehat{y}_i = c|x_i) - \max_{c\in\Omega/c^+} p(\widehat{y}_i = c|x_i) \tag{1}$$

where $\Omega = (1, 2, \cdots, C)$; $c^+ = \arg\max_{c\in\Omega}(p(\widehat{y}_i = c|x_i))$ represents the class to which x_i is most likely to belong; Ω/c^+ denotes the set of all the class labels in Ω except c^+. If $\max_{c\in\Omega} p(\widehat{y}_i = c|x_i) \gg \max_{c\in\Omega/c^+} p(\widehat{y}_i = c|x_i)$, the probability that x_i belongs to class c^+ is high. If $\max_{c\in\Omega} p(\widehat{y}_i = c|x_i)$ and $\max_{c\in\Omega/c^+} p(\widehat{y}_i = c|x_i)$ are close, x_i is more likely to be misclassified. Therefore, selecting pseudo-labelled target samples with top soft confidence scores can prevent adding the mislabelled samples to the training set. It is noticed that unlike other works [26, 28] using hard confidence, i.e., $U_i = \max_{c\in\Omega} p(\widehat{y}_i = c|x_i)$, the relative confidence is more reasonable and better. At the beginning of training, the classification ability of the network is weak so that it cannot get a confident and reliable output over categories. For example, we think that a sample with the output of (0.2, 0.6, 0.2, 0.1) is more reliable than a sample with the output of (0.65, 0.35, 0, 0) although the latter gets higher confidence in the first class. In real-world applications, the relative confidence can be used on a probabilistic model rather than a decision model, since it is calculated by the two highest probabilities from the model's prediction. Not just neural networks, other machine learning methods, such as SVMs and Regression, can determine the relative confidence through subtracting the second-highest category probability from the highest probability.

We adopt the following selective pseudo-labelling strategy. In the k - th iteration, the relative confidence of all the target domain samples is calculated. For class c, the number of selected pseudo-labelled target domain samples $N(c, k)$ is determined by

$$N(c, k) = \min\{N_c \times k/T, n_t(c, k)\} \tag{2}$$

where T is the number of iterations.

N_c is the average sample number of C classes in the target domain.

$n_t(c,k)$ is the number of the target domain samples which are classified into the c - th class in the k - th iteration. Our pseudo-labelling selection allows balanced pseudo-labelled target samples across different classes. The number of predicted pseudo-labels $n_t(c,k)$ increases as the iteration proceeds. As a result, there can be a large number of selected pseudo-labelled samples for 'easy' class while very limited pseudo-labelled samples for other samples. We make a minimization with the $N_c \times k/T$ in order to prevent the network to be biased to the 'easy' class so that pseudo-labelled target samples will contribute to the alignment of distribution for each class during learning.

3.3 Loss function of class confusion

After the pseudo-labelled target domain samples are added to the training set, the classifier can directly learn knowledge from the target domain. However, the accumulated error caused by mis-labelling still exists. According to Eq. (1), in each iteration the target domain samples with high relative confidence are more likely to be selected as the training samples in the next iteration than those with relative confidence scores. Obviously, samples that make the classifier ambiguous and unconfident across classes may not be fully utilized in domain adaptation learning. As shown in Fig. 4, the selected target domain samples correspond to a sparse classification confusion matrix, while the classification confusion matrix of the unselected target domain samples is more dispersed.

Minimum class confusion (MCC) [14] is a general loss function which can be characterized as a domain adaptation method without explicitly deploying domain alignment since it only uses the target domain data. When the selected target domain samples are added to the training set for the next iteration, the unselected samples will be used to calculate the minimum class confusion as the regularization item for classifier training. Introducing MCC has the following advantages. Firstly, it can be used as a general regularization that prevents the network from being stuck in the local optima. Secondly, in each iteration, the samples farthest to the classification hyperplane are selected, as shown in Fig. 4. Utilizing class confusion term of the unselected target domain samples as a part of loss function in current iteration makes complementary improvement from the target domain. After current iteration, they are more likely to be given higher certainty and selected in the next

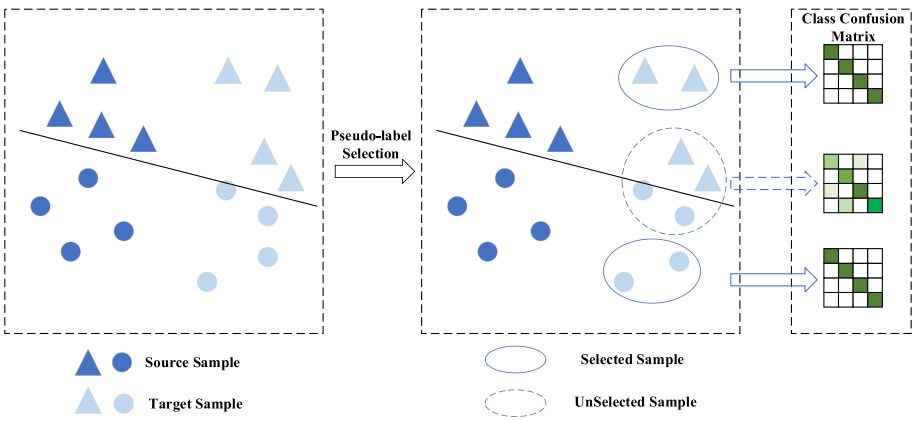

Fig. 4 Classification confusion matrices of selected and unselected samples in the target domain

iteration for class prediction. Thirdly, MCC can largely accelerate convergence and achieve a high domain adaptation performance with limited number of iterations.

We add class confusion as a regularization on unlabelled target domain samples. The confusion between different classes can be naturally described by an inner-product between the classifier predictions and their responses. Firstly, temperature rescaling [30] is added to the softmax output of the classifier to alleviate the overconfident predictions. The probability that the i-th instance belongs to the j-th class is expressed as

$$Z_{i,j} = \frac{\exp(Y_{i,j}/T)}{\sum_{j'=1}^{C} \exp(Y_{i,j}/T)} \tag{3}$$

where $\widehat{Y}_{i,j}$ is the logit produced by the network of the i-th instance. The class correlation between classes j and j' is defined as

$$C_{jj'} = \hat{z}_j^T \hat{z}_{j'} \tag{4}$$

where $\widehat{z}_j$ denotes the probabilities that the samples in each batch come from the j-th class. The class correlation measures the probability that the classifier simultaneously classifies the examples into the j-th and j'-th class.

Those examples with higher certainty in class predictions given by the classifier are more reliable and should contribute more to the pairwise class confusion [14]. Furthermore, a weighting mechanism based on the uncertainty is added such that class confusion could highlight the samples with higher certainty in class predictions and ignore the samples that shows little category tendency. The entropy function is used as the measure of uncertainty, which is defined as

$$H(\hat{z}_i) = -\sum_{j=1}^{C} Z_{i,j} \log Z_{i,j} \tag{5}$$

With weighting mechanism, preliminary definition of class confusion is defined as

$$C_{jj'} = \hat{z}_j^T W \hat{z}_{j'} \tag{6}$$

where

$$W_{ii} = \frac{B(1 + \exp(-H(\hat{z}_i)))}{\sum_{i'=1}^{B} 1 + \exp(-H(\hat{z}_i))} \tag{7}$$

In Eq. (7), W_{ii} is the probability of quantifying the importance of the i-th sample for modeling the class confusion. W is the corresponding diagonal matrix in Eq. (6). Finally, the formal class confusion loss function that is native for the mini-batch SGD optimization is written as

$$L_{CC}(\widehat{Y}) = \frac{1}{C}\sum_{j=1}^{|C|}\sum_{j'\neq j}^{|C|}\left|\frac{C_{jj'}}{\sum_{j''=1}^{C} C_{jj''}}\right| \tag{8}$$

In Eq. (8), it is noted that a category normalization technique [31] is adopted to prevent a severe class imbalance when the number of classes is large.

3.4 Total loss function

The total loss function of the classifier L_{total} consists of two parts, the classification loss L_{cls} and the class confusion loss L_{CC}.

$$L_{\text{total}} = L_{\text{cls}}\left(X_s, X_t^{\text{pse}}\right) + \lambda L_{CC}\left(X_t^{\text{unselected}}\right) \tag{9}$$

where X_s refers to the source domain samples.X_t^{pse} the pseudo-labelled target domain samples in current iteration.$X_t^{\text{unselected}}$ the unselected target domain samples. L_{cls} is given by

$$L_{\text{cls}}\left(X_s, X_t^{\text{pse}}\right) = -\frac{1}{N_s}\sum_{i}\sum_{c=1}^{C} y_{ic}\log\left(p_{ic}\right) - \frac{1}{N_t^{\text{pse}}}\sum_{j}\sum_{c=1}^{C} y_{jc}^{\text{pse}}\log\left(p_{jc}\right) \tag{10}$$

where N_s and N_t^{pse} are the numbers of the source domain samples and the pseudo-labelled target domain samples, respectively. y_{ic} and y_{jc}^{pse} represent the sign functions of the i-th source domain sample and the j-th pseudo-labelled target domain sample, respectively, as given by Eq. (11) and Eq. (12). p_{ic} and p_{jc} refers to the probabilities that the i-th source domain sample and the j-th pseudo-labelled target domain sample belongs to class c.

$$y_{ic} = \begin{cases} 1 & \text{if i} = \text{c} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$y_{jc}^{\text{pse}} = \begin{cases} 1 & \text{if i} = \text{c} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

3.5 Algorithm of the proposed SPL-CCR for SAR target recognition

To summarize, the proposed method of SPL-CCR for SAR target recognition is shown in Algorithm 1. In the beginning, the training set initialized as the source domain data X_s is used to train the classifier G. Next, in each iteration, the samples with the highest probability to be correctly classified from the unlabelled target domain are assigned with pseudo-labels and selected. The selected samples are used to form new training set and retrain the classifier G. The process is repeated until the iteration condition is reached or the pseudo-labels no longer change.

4 Results and discussion

In this section, we describe our experiments on the MSTAR dataset for SPL-CCR for SAR image classification. Our method is firstly compared with state-of-the-art unsupervised domain adaptation methods to evaluate its effectiveness. Then, we use the t-SNE technique to visualize the extracted features from the source and target domains. Besides, we quantitatively measure the domain discrepancy across different methods using $\mathcal{A}$-distance [32]. Finally, we investigate the performance under conditions with limited source and target samples.

4.1 Dataset description

The MSTAR dataset [33] contains SAR images of ten targets including tanks, armored vehicles, weapon systems and military engineer vehicles (armored personal carrier: BMP-2, BRDM-2, BTR-60, and BTR-70; tank: T-62, T-72; weapon system: 2S1; air defense unit: ZSU-234; truck: ZIL-131; bulldozer: D7). The data were collected with a Sandia X-band radar. The range and cross-range resolution are identical and equal to 0.30 m.

Considering the sensitivity of SAR images to depression angles, we conduct the experiments on images with different depression angles to explore the cross-domain classification capability of our method. The MSTAR dataset contains SAR images with four depression angles, i.e., 15°,17°, 30° and 45°. There are four classes (2S1, BRDM2, T72 and ZSU234) of target images covers all these depression angles. Figure 5 shows some optical and SAR images of the four classes of targets with different depression angles. The numbers of SAR images of different targets under different depression angles are given in Table 2. We construct six domain adaptation tasks by setting different source and target data configurations, i.e., 17° → 30°, 30° → 17°, 17° → 45°, 45° → 17°, 30° → 45° and 45° → 30°. Note that data under depression angle of 15° is not involved in our experiment since it shows little difference with data under depression angle of 17°.

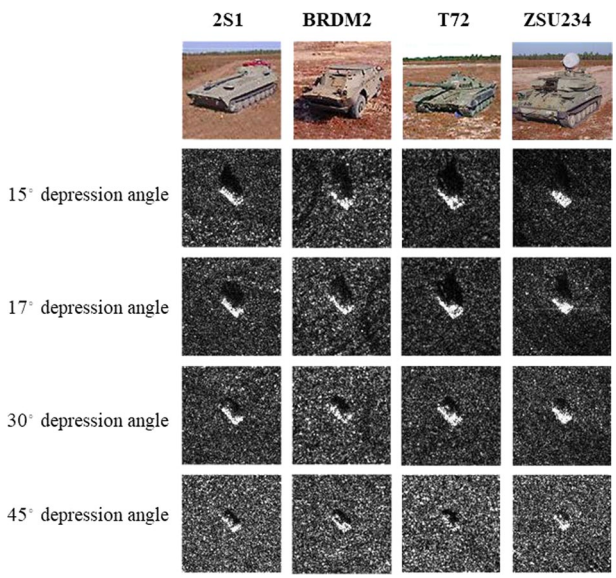

Fig. 5 Optical and SAR images of four classes of targets in the MSTAR dataset

Table 2 Numbers of SAR images of different targets under different depression angles

Class	Type	Depression Angle			
		17°	15°	30°	45°
2S1	B01	299	274	288	303
BRDM2	E-71	298	274	287	303
ZSU234	d08	299	274	288	303
T72	SN_132, A64	232	196	288	303

4.2 Experimental setting

Each sample in the MSTAR dataset is cropped to the size of 128×128 pixels and no image augmentation and pre-processing algorithm is applied. The algorithms are implemented in Pytorch1.7. The classification model is trained iteratively by a stochastic gradient descent (SGD) optimization algorithm with a momentum of 0.9. The learning rate is adjusted using the simulated annealing strategy with the following schedule

$$\eta_p = \eta_0(1 + \alpha p)^{-\beta} \tag{13}$$

where η_0 is the initial learning rate.$\alpha = 0.001$.$\beta = 0.75$. p represents the ratio of current epoch and total epochs, gradually increasing from 0 to 1.

The pseudo-labels of the target domain samples in the last epoch are used as the predictions of the classification model. Different from most traditional domain adaptation methods, pseudo-labelling is implemented in an iterative way. Therefore, the epoch number in each iteration for pseudo-labelling is set as 10 and the maximum iteration number is set as 10. Epoch number for other domain adaptation methods is set as 100, for a fair comparison on training total epochs with methods based on pseudo-labelling.

4.3 Comparison with other approaches

We compare our method with the most competitive classification methods including the supervised-learning-based (ResNet18) and unsupervised domain adaptation methods (SPL, DDC, DaNN, CDAN, MCC, JAN, MCD). SPL refers to the selective pseudo-labelling method without class confusion regularization. All hyper-parameters of the compared method are adopted in the same scheme for experimental fairness. We apply all methods based on Resnet-18 to evaluate their performance. The average classification accuracy of each method is reported on three random experiments and each method is trained for 80 epochs per experiment. The batch size is set to 8. For MMD-based methods, i.e., DDC and JAN, we adopt a Gaussian kernel with bandwidth set to median pairwise squared distances. Each method is optimized using SGD with a momentum of 0.9 and a weight decay of 5×10^{-4}, under the same learning rate adjusting scheme with SPL and SPL-CCR. The classification results of six domain adaptation tasks with different methods are given in Table 3. The classification precision, namely the ratio of the numbers of correctly classified samples and total samples, is chosen as the classification evaluation metric in the following tables.

According to Table 3, we find that the differences of depression angles result in large performance differences in classification. We should first emphasize that all methods

Table 3 Classification results of different methods on MSTAR dataset

	17°→ 30°	**30°→ 17°**	**17°→ 45°**	**45°→ 17°**	**30°→ 45°**	**45°→ 30°**	**Average**
ResNet18[27]	96.48	99.91	50.58	55.59	61.63	69.24	72.24
DDC [8]	95.48	98.89	65.84	74.38	75.04	82.19	81.97
DaNN [20]	99.31	**100.00**	66.79	77.79	90.10	**99.74**	88.95
CDAN [12]	**100.00**	**100.00**	71.87	86.52	93.23	98.22	91.64
MCC [14]	99.39	99.96	82.10	69.24	90.22	83.97	87.48
JAN [11]	99.22	99.51	25.58	50.40	43.81	55.78	62.38
MCD [13]	88.92	98.40	46.41	64.63	48.93	69.50	69.47
SPL	99.52	**100.00**	84.28	87.06	90.47	97.70	93.17
SPL-CCR	99.96	**100.00**	**91.87**	**91.98**	**96.49**	99.04	**96.56**

have achieved above 99% recognition accuracy over the source domain due to the strong fitting ability of the ResNet-18 network, which is not shown in the table. However, for all the methods, the classification performances decrease with the increasement of the depression angle difference. Our SPL method achieves an average accuracy of 93.17% over the six domain adaptation tasks which outperforms all the comparative methods except SPL-CCR. This proves that generating pseudo-labels can directly capture the information of the target domain data. Although the feature distributions are not strictly aligned in the hidden space of the middle layers of the network, the SPL method is able to comprehensively represent the target domain data. As for the proposed SPL-CCR method, its average accuracy over the six tasks increases to 96.56%. On the one hand, class confusion regularization is used as a domain adaptation loss function, which can narrow the distribution discrepancy of the output between the source and target domains. Thus, the network will generate more accurate pseudo-labels in the next iteration. On the other hand, class confusion regularization improves the convergence of the method. Higher domain adaptation performance can be achieved under limited number of iterations.

In the task of 17° → 45°, SPL-CCR, SPL and MCC reach the top three scores of classification accuracy, which are 91.87%, 84.28% and 82.10%, respectively. The accuracy of SPL-CCR is higher than that of MCC by 9.77%. In the task of 45° → 17°, SPL-CCR, SPL and CDAN achieve the top three scores of classification accuracy, which are 91.98%, 87.06% and 86.52%, respectively. The accuracy of SPL-CCR is higher than that of CDAN by 5.46%. It is noted that JAN and CDAN have a lower performance in the tasks of 17° → 45° and 45° → 17°, compared to other experimental setups. Especially, JAN performs worst compared to other methods. Due to huge domain discrepancy, hard-to-transfer examples with uncertain predictions may deteriorate the conditional adversarial adaptation procedure. Hence, features cannot be aligned through capturing the cross-covariance of feature representation and classification prediction. In the task of 30° → 45°, SPL-CCR, CDAN and SPL achieve the top three scores of classification accuracy, which are 96.49%, 93.23% and 90.47, respectively. CDAN outperforms SPL, but its accuracy is 3.26% lower than that of SPL-CCR. In other three tasks, our SPL-CCR method does not significantly outperform the other methods. This indicates that feature-based domain adaptation methods are sufficient to improve model generalization when the difference of depression angles is small. However, if the difference

of depression angles is large, e.g., the 17° → 45° or 45° → 17° tasks in our experiment, classic unsupervised domain adaptation methods fail to align the feature distributions well, due to large domain discrepancy. Our SPL-CCR method can transfer the knowledge of the source domain to the target domain by iteratively generating pseudo-labels and performs better even if the difference of depression angles is large.

4.4 Feature visualization

For qualitative comparison of different methods, we use the t-SNE technique to visualize the features of the source domain and target domain over the tasks of 17° → 30°, 17° → 45° and 30° → 45°. Our SPL-CCR method are compared with ResNet18, DDC, DaNN and CDAN, since they represent different types of methods and most of them achieve the highest accuracy in a certain task.

Firstly, we visualize the features in the domain adaptation layer of each network. Here, the domain adaptation layer refers to the layer before the output layer. Figure 6 shows the visualization results. Blue and red dots represent the source domain samples and target domain samples, respectively. Figure 6a shows the feature distribution of the domain adaptation layer of ResNet18. The network is trained using only the source domain data. Misalignment of the source domain and target domain samples are quite severe, which leads to bad classification performance. According to Fig. 6b–d, domain adaptation methods such as DDC, DaNN and CDAN can align samples from the source domain and target domain over the tasks of 17° → 30° and 30° → 45°. However, we can see that the source domain samples and target domain samples are not well aligned in the task of 17° → 45°. This indicates that the domain adaptation loss usually plays a regularization role in the training process and improves the classification performance in the target domain by reduce domain discrepancy. But, it deteriorates the classification in the source domain and results in the nonnegligible generalization error in the target domain. As for our SPL-CCR method, according to Fig. 6e, since the proposed method does not use metric criteria or adversarial learning methods for domain adaptation, the distribution differences still exist. However,

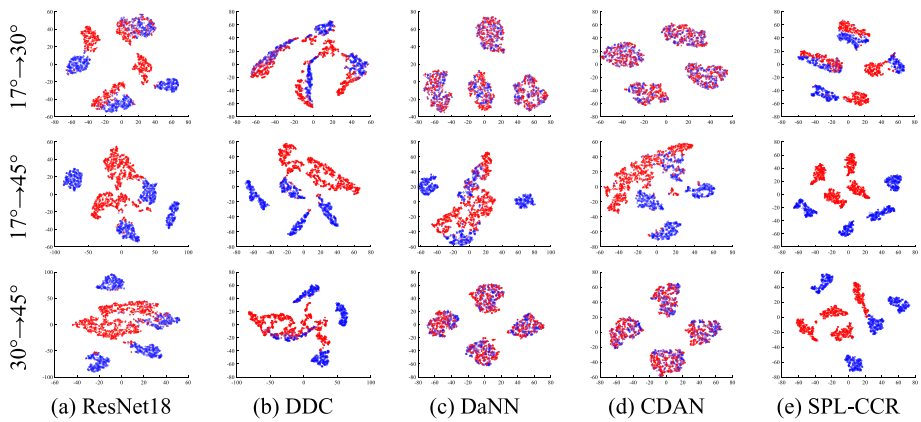

Fig. 6 Visualization of features in the domain adaptation layer of each network (blue and red dots represent the source domain samples and target domain samples, respectively). **a** ResNet18. **b** DDC. **c** DaNN. **d** CDAN. **e** SPL-CCR. Each row corresponds to a single task, i.e., 17° → 30°, 17° → 45° and 30° → 45°

both the source and target domain data show strong separability, which indicates our method is able to capture the data modality of the target domain and adapt the data distribution.

Secondly, we visualize the feature in the output layer of each network, as shown in Fig. 7. The feature distribution in the output layer of a network can directly reflect the generalization error in the source and target domains as well as domain adaptation ability. According to Fig. 7, our SPL-CCR method can extract features which are both domain and class discriminative over the three tasks. It should be pointed out that other domain adaptation methods such as DaNN and CDAN can also achieve good adaptation ability in the tasks of $17° \rightarrow 30°$ and $30° \rightarrow 45°$. Especially in the task of $30° \rightarrow 45°$, the source domain and target domain samples are better aligned by DaNN or CDAN than SPL-CCR. However, in the task of $17° \rightarrow 45°$, SPL-CCR shows stronger ability in feature alignment than DaNN and CDAN. On the contrary, although our SPL-CCR method does not perfectly align the features from the source and target domains, each class is still distinguishable in both domains over the three tasks.

4.5 Domain discrepancy comparison

$\mathcal{A}$-distance is often used to measure domain discrepancy in domain adaptation researches. It is defined as

$$d_{\mathcal{A}}(D_s, D_t) = 2(1 - 2L_g) \tag{14}$$

where D_s and D_t denote the source and target domain samples, respectively. L_g is the generalization error of a two-sample classification. Here we use a single-layer network and sigmoid function as a binary classifier.

The results of $\mathcal{A}$-distance of different methods over all the six tasks ($17° \rightarrow 30°$, $30° \rightarrow 17°$, $17° \rightarrow 45°$, $45° \rightarrow 17°$, $30° \rightarrow 45°$ and $45° \rightarrow 30°$) are shown in Fig. 8. As we can see, the $\mathcal{A}$-distance of our SPL-CCR is significantly lower than those of other methods. This indicates that class confusion regularization and pseudo-labelling can reduce cross-domain divergence more effectively.

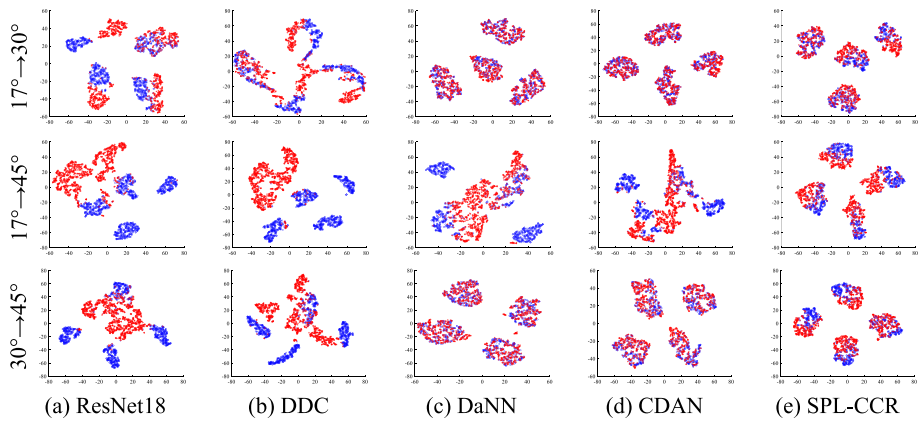

Fig. 7 Visualization of features in the output layer of each network (blue and red dots represent the source domain samples and target domain samples, respectively). **a** ResNet18. **b** DDC. **c** DaNN. **d** CDAN. **e** SPL-CCR. Each row corresponds to a single task, i.e., $17° \rightarrow 30°$, $17° \rightarrow 45°$ and $30° \rightarrow 45°$

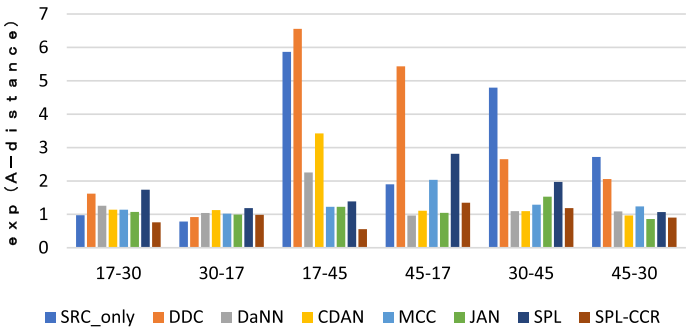

Fig. 8 The results of $\mathcal{A}$-distance of different methods over all the six tasks

4.6 Classification with limited samples

This subsection compares our SPL-CCR method with other methods in the classification performance under conditions of limited samples. Since the training data includes both the source and target domain data, N_t samples are randomly selected for each class as the training data in both domains. N_t is set to 10, 20 and 30. The test data are all samples of the target domain. Numbers of training and test samples in each task are given in Table 4.

Tables 5, 6 and 7 show the classification results of different methods over all the six tasks with limited samples, i.e. 10, 20, and 30 samples per class, are randomly selected from the training set in the source domain and target domain, respectively. It can be obviously seen that as the number of training samples increases, the classification

Table 4 Experimental data for classification with limited samples

	Number of training samples in the source domain	Number of training samples in the target domain	Number of test samples
17° → 30°	10/20/30	10/20/30	1151
30° → 17°	10/20/30	10/20/30	1128
17° → 45°	10/20/30	10/20/30	1212
45° → 17°	10/20/30	10/20/30	1128
30° → 45°	10/20/30	10/20/30	1212
45° → 30°	10/20/30	10/20/30	1151

Table 5 Classification results of different methods with 10 samples randomly selected for each class

	17° → 30°	**30° → 17°**	**17° → 45°**	**45° → 17°**	**30° → 45°**	**45° → 30°**	**Average**
ResNet18[27]	82.11	87.94	43.61	36.26	42.16	37.66	54.96
DDC[8]	67.20	80.45	43.28	59.53	52.81	50.87	59.02
DaNN[20]	92.66	89.32	60.19	62.72	65.31	55.30	70.92
CDAN[12]	89.92	93.93	50.33	59.57	57.43	58.82	68.33
MCC[14]	85.23	92.46	59.86	57.36	**65.88**	49.74	68.42
JAN[11]	87.79	93.66	42.41	36.26	58.62	49.44	61.36
MCD[13]	52.48	44.95	41.67	33.25	41.60	39.53	42.24
SPL	**93.66**	**96.14**	61.96	69.33	61.74	59.25	73.68
SPL-CCR	88.09	95.79	**63.04**	**74.48**	62.50	**61.38**	**74.21**

Table 6 Classification results of different methods with 20 samples randomly selected for each class

	17°→30°	**30°→17°**	**17°→45°**	**45°→17°**	**30°→45°**	**45°→30°**	**Average**
ResNet18[27]	88.40	83.11	42.52	53.19	55.53	53.78	62.76
DDC[8]	84.80	80.05	55.28	70.52	55.57	58.82	67.51
DaNN[20]	91.14	92.95	66.30	64.63	69.14	**68.03**	75.36
CDAN[12]	86.66	94.28	69.06	60.99	71.04	65.86	74.65
MCC[14]	91.14	93.44	66.67	49.03	66.58	42.57	68.24
JAN[11]	90.27	82.14	38.57	47.74	52.68	46.65	59.67
MCD[13]	68.42	49.96	33.42	34.00	39.74	36.50	43.67
SPL	91.31	96.28	**75.91**	**64.50**	71.45	66.99	77.74
SPL-CCR	**93.75**	**96.51**	73.40	63.45	**77.00**	64.51	**78.10**

Table 7 Classification results of different methods with 30 samples randomly selected for each class

	17°→30°	**30°→17°**	**17°→45°**	**45°→17°**	**30°→45°**	**45°→30°**	**Average**
ResNet18[27]	91.40	90.82	52.06	53.99	61.76	60.30	68.39
DDC[8]	86.58	89.85	56.81	76.15	61.80	61.88	72.18
DaNN[20]	92.53	97.03	69.80	67.20	70.42	75.24	78.70
CDAN[12]	95.83	97.61	60.68	72.34	68.94	77.24	78.77
MCC[14]	95.79	98.27	64.03	72.21	72.32	83.06	80.94
JAN[11]	94.05	94.37	39.48	42.51	55.07	56.95	63.74
MCD[13]	70.68	31.74	35.50	30.67	43.30	31.49	40.56
SPL	95.09	97.83	65.68	78.37	**87.88**	**81.84**	84.45
SPL-CCR	**95.66**	**98.89**	**77.48**	**85.77**	80.73	78.24	**86.13**

accuracy is improved since more source domain samples with different imaging environments and target poses are added in the training process.

The classification results with limited samples are consistent with the results in Sect. 4.3, where the training samples are more sufficient. Although the improvement is reduced, our SPL-CCR method still outperforms other methods and proves its superiority under limited-sample conditions. It also suggests that domain adaptation methods based on pseudo-labelling are effective in SAR target recognition with sample limitation since pseudo-label can make a label complement to prevent network overfitting.

With very limited samples, SPL-CCR or SPL do not achieve the best performance over some tasks, e.g., 30°→45° with 10 samples selected for per class and 45°→30° with 20 samples selected for per class. The reason may be that mis-labelling issue is more severe in the limited sample cases, which causes error accumulation and classification performance drop. Furthermore, Tables 5, 6 and 7, combination of class confusion regularization with SPL displays noticeable accuracy enhancements in most tasks, which proves the effectiveness of class confusion regularization under limited-sample conditions.

5 Conclusions

In this paper, we proposed a novel method for SAR target classification from a perspective of domain adaptation to tackle performance degradation problem caused by variant imaging conditions. A selective pseudo-labelling strategy based on the BT

criterion and class confusion regularization is designed. Part of the target domain samples are assigned pseudo-labels and added to the training set in an iterative way. Therefore, data information in the target domain can be directly studied. Considering the problem of error accumulation of pseudo-labelling, class confusion loss is introduced into the iterative process as a regularization term, which improves the network's adaptation to the target domain samples. We conducted the experiments on SAR images with different depression angles to explore the cross-domain classification capability of our method. Based on the MSTAR dataset, six configurations of depression angles are considered to construct the source domain and target domain data for target classification. The proposed SPL-CCR method achieved an average accuracy of 96.56% over all the six tasks, which is significantly higher than those of other comparative methods such as ResNet18, DDC, DaNN, CDAN, MCC, JAN and MCD. The t-SNE feature visualization results show that the proposed method has strong ability in feature alignment across two domains and extracts features maintaining good separability from the target domain. Besides, our method also shows its superiority under limited-sample conditions. At present, our work mainly focuses on the variation of depression angle. In the future, we will further study SAR ATR tasks with other type of imaging condition variations.

Abbreviations

SAR	Synthetic aperture radar
BT	Breaking ties
MSTAR	Moving and stationary target acquisition and recognition
DL	Deep learning
ATR	Automatic target recognition
CNN	Convolutional neural network
SVM	Support-vector machine
DDC	Deep domain confusion
DAN	Deep adaptation networks
DaNN	Domain adversarial neural network
CDAN	Conditional domain adaptation network
MCC	Minimum class confusion
JAN	Joint adaptation networks
MCD	Maximum classifier discrepancy
t-SNE	T-distributed stochastic neighbor embedding
RKHS	Reproducing kernel Hilbert space
SPL-CCR	Selective pseudo-labelling with class confusion regularization
SPL	Selective pseudo-labelling
SGD	Stochastic gradient descent

Acknowledgements
The authors would like to thank the handing Associate Editor and the anonymous reviewers for their valuable comments and suggestions for this paper.

Author contributions
Lingjun Zhao designed the work, analyzed and interpreted the data, and drafted the manuscript. Qishan He participated in the design of the study, performed the experiments and analysis, and helped to draft the manuscript. Ding Ding and Siqian Zhang contributed to literature investigation. Gangyao Kuang and Li Liu contributed to revise the manuscript. All the authors read and approved the final manuscript.

Funding
This work was supported by the Natural Science Foundation of Hunan Province, China under Grant 2021JJ30780.

Availability of data and materials
Please contact the authors for data requests.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 1 April 2022 Accepted: 24 August 2022
Published online: 11 September 2022

References

1. S. Chen, H. Wang, F. Xu, Y. Jin, Target classification using the deep convolutional networks for SAR images. IEEE Trans. Geosci. Remote Sens. **54**(8), 4806–4817 (2016)
2. S.A. Wagner, SAR ATR by a combination of convolutional neural network and support vector machines. IEEE Trans. Aerosp. Electron. Syst. **52**(6), 2861–2872 (2016)
3. Z. Lin, K. Ji, M. Kang, X. Leng, H. Zou, Deep convolutional highway unit network for SAR target classification with limited labeled training data. IEEE Geosci. Remote Sens. Lett. **14**(7), 1091–1095 (2017)
4. R. Qin, X. Fu, J. Dong, W. Jiang, A semi-greedy neural network CAE-HL-CNN for SAR target recognition with limited training data. Int. J. Remote Sens. **41**(20), 7889–7911 (2020)
5. M. Rostami, S. Kolouri, E. Eaton, K. Kim, Deep transfer learning for few-shot SAR image classification. Remote Sens. **11**(11), 1374 (2019)
6. Y. Ma, Y. Liang, W. Zhang, S. Yan, SAR target recognition based on transfer learning and data augmentation with LSGANs. Paper presented at 2019 Chinese Automation Congress (CAC), HangZhou, China, 22–24 Nov, (2019)
7. Q. He, L. Zhao, K. Ji, G. Kuang, SAR target recognition based on task-driven domain adaptation using simulated data. IEEE Geosci. Remote Sens. Lett. **19**, 1–5 (2022)
8. E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: maximizing for domain invariance, Computer Science, (2014)
9. M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks. Paper presented at 2015 International Conference on Machine Learning (ICML), Miami, Florida, USA, 9–11 Dec, (2015)
10. B. Sun, K. Saenko, Deep Coral: Correlation alignment for deep domain adaptation. Paper presented at 2016 European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 Oct, (2016)
11. M. Long, H. Zhu, J. Wang, M.I. Jordan, Deep transfer learning with joint adaptation networks. Paper presented at 2017 International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 Aug, (2017)
12. M. Long, Z. Cao, J. Wang, M.I. Jordan, Conditional adversarial domain adaptation. Paper presented at 2018 Neural Information Processing Systems (NIPS), Montréal, Canada, 03–06 Dec, (2018)
13. K. Saito, K. Watanabe, Y. Ushiku, T. Harada, Maximum classifier discrepancy for unsupervised domain adaptation. Paper presented at 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, Utah, USA, 18–22 June, (2018)
14. Y. Jin, X. Wang, M. Long, J. Wang, Minimum class confusion for versatile domain adaptation. Paper presented at 2020 European Conference on Computer Vision (ECCV), 23–28 Aug, (2020)
15. W. Deng, Q. Liao, L. Zhao, D. Guo, G. Kuang, D. Hu, L. Liu, Joint clustering and discriminative feature alignment for unsupervised domain adaptation. IEEE Trans. Image Process. **30**, 7842–7855 (2021)
16. E. Schonfeld, S. Ebrahimi, S. Sinha, T. Darrell, Z. Akata, Generalized zero-and few-shot learning via aligned variational autoencoders. Paper presented at IEEE Computer Vision and Pattern Recognition (CVPR), Long Beach, California, USA, 16–20 June, (2019)
17. L. Van der Maaten, G. Hinton, Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11), (2008)
18. S. Ben-David, J. Blitzer, K. Crammer, F. Pereira, Analysis of representations for domain adaptation. Paper presented at Neural Information Processing Systems (NIPs), (2007)
19. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Paper presented at 2014 Neural Information Processing Systems (NIPs), Montreal, Canada, (2014)
20. Y. Ganin, V. Lempitsky, Unsupervised domain adaptation by backpropagation. Paper presented at 2015 International Conference on Machine Learning (ICML), Lile, France, 6–11 July, (2015)
21. X. Zhu, A.B. Goldberg, Introduction to semi-supervised learning. Synth. Lectures Artif. Intell. Mach. Learn. **3**(1), 1–130 (2009)
22. X. Li, Q. Sun, Y. Liu, Q. Zhou, S. Zheng, T.-S. Chua, B. Schiele, Learning to self-train for semi-supervised few-shot classification. Paper presented at 2019 Neural Information Processing Systems (NIPs), Jaipur, India, (2019)
23. M. Long, J. Wang, G. Ding, J. Sun, P.S. Yu, Transfer Feature Learning with Joint Distribution Adaptation. Paper presented at 2013 IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 3–6 Dec, (2013)
24. J. Zhang, W. Li, P. Ogunbona, Joint geometrical and statistical alignment for visual domain adaptation. Paper presented at 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 21–26 July, (2017)
25. Z. Pei, Z. Cao, M. Long, J. Wang, Multi-adversarial domain adaptation. Paper presented at 2018 AAAI Conference on Artificial Intelligence (AAAI), New Orleans, Louisiana, USA, 2–7 Feb, (2018)
26. C. Chen, W. Xie, W. Huang, Y. Rong, X. Ding, Y. Huang, T. Xu, J. Huang, Progressive feature alignment for unsupervised domain adaptation. Paper presented at 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, California, USA, 16–20 June, (2019)
27. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. Paper presented at IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 26 June–1 July, (2016)
28. Q. Wang, T. Breckon, Unsupervised Domain Adaptation via Structured Prediction Based Selective Pseudo-labeling. Paper presented at 2020 AAAI Conference on Artificial Intelligence, Hilton New York Midtown, New York, USA, 7–12, Feb, (2020)
29. B. Demir, C. Persello, L. Bruzzone, Batch-mode active-learning methods for the interactive classification of remote sensing images. IEEE Trans. Geosci. Remote Sens. **49**(3), 1014–1031 (2010)

30. R. Min, H. Lan, Z. Cao, Z. Cui, A gradually distilled CNN for SAR target recognition. IEEE Access **7**, 42190–42200 (2019)
31. U. Von Luxburg, A tutorial on spectral clustering. Stat. Comput. **17**(4), 395–416 (2007)
32. S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J.W. Vaughan, A theory of learning from different domains. Mach. Learn. **79**(1), 151–175 (2010)
33. T.D. Ross, S.W. Worrell, V.J. Velten, J.C. Mossing, M.L. Bryant, Standard SAR ATR Evaluation Experiments Using the MSTAR Public Release Data Set. Paper presented at Algorithms for Synthetic Aperture Radar Imagery, (1998)

Publisher's Note