Discriminative Adaptation Regularization Framework-Based Transfer Learning for Ship Classification in SAR Images

Yongjie Xu, *Student Member, IEEE*, Haitao Lang, *Member, IEEE*, Lihui Niu, and Chenguang Ge

***Abstract*— Ship classification in synthetic-aperture radar (SAR) images is of great significance for dealing with various marine matters. Although traditional supervised learning methods have recently achieved dramatic successes, but they are limited by the insufficient labeled training data. This letter presents a novel unsupervised domain adaptation (DA) method, termed as discriminative adaptation regularization framework-based transfer learning (D-ARTL), to address the problem in case that there is no labeled training data available at all in the SAR image domain, i.e., target domain (TD). D-ARTL improves the original ARTL by adding a novel source discriminative information preservation (SDIP) regularization term. This improvement achieves an efficient transfer of interclass discriminative ability from source domain (SD) to TD, while achieving the alignment of cross-domain distributions. Extensive experiments have verified that D-ARTL outperforms state-of-the-art methods on the task of ship classification in SAR images by transferring the automatic identification system (AIS) information.**

***Index Terms*— Ship classification, synthetic-aperture radar (SAR), transfer learning, unsupervised domain adaptation.**

I. Introduction

SATELLITE exploration technology based on synthetic-aperture radar (SAR) has been widely used in many marine matters. Accurate recognition of ship type via SAR images is of great significance for many missions, such as combating illegal ship invasion, participating in sea rescue, and managing maritime traffic. Ship classification in SAR images based on supervised learning methods has recently attracted great interest. Ji *et al.* [1] proposed to combine support vector machine (SVM) and discriminative information in multiple classifiers to achieve the complementary advantages between classifiers. Xing *et al.* [2] proposed a dictionary based on sparse features (ship length, ship width, and scattering features), which further improved the reliability of the dictionary while reducing the calculation burden. Lang *et al.* [3] proposed the method of joint feature and classifier selection by integrating the classifier selection strategy into a wrapper feature selection framework. Lang and Wu [4] proposed to improve the ship classification performance by fusing the naive geometric features (NGFs) via multiple kernel learning (MKL) algorithm. Bentes *et al.* [5] proposed a multiple-resolution convolution neural networks (MR-CNNs) for the first time to conduct maritime target classification in SAR images.

Although supervised learning approaches have recently achieved dramatic successes on the task of ship classification in SAR images, those methods are depended on the availability of massive labeled training data, especially for the deep learning method that needs to train complex deep neural networks. Unfortunately, obtaining labels for ship in SAR images is often a time-consuming and costly process that requires human experts and expensive ground campaigns, and is not even possible in many real-world application scenario. Domain adaptation (DA) is a promising alternative to solve the problem when it is difficult to obtain sufficient labeled data [6]. The main idea of DA is to use the knowledge of a source domain (SD) that is easy to obtain labeled data to assist in the classification in target domain (TD). DA has improved the state-of-the-art in many classification tasks, including optical, infrared, or medical images, but its use in SAR classification problem has been less exploited. To the best of our knowledge, Lang *et al.* [7] proposed to improve the performance of ship classification in SAR image domain by transferring the knowledge of automatic identification system (AIS) domain. They proved that when a few training samples in TD are available, the performance of traditional supervised classification can be significantly improved by transferring AIS knowledge.

Following the idea in [7], this letter also focuses on how to utilize AIS data (SD) to improve ship classification task in SAR domain (TD). The main difference with [7] is, this letter investigates an extreme case that there is no labeled data in the TD at all. This is a compelling problem and the main challenge comes from two aspects: 1) since there is no labeled sample, it needs to develop an unsupervised DA approach and 2) as shown in Fig. 1, since the probability distribution in AIS (SD) and SAR (TD) domains changes tremendously and has very different statistical properties, it needs to develop an effective strategy to address the domain shift problem. Recent works aim to discover a good feature representation across domains, which can simultaneously reduce the

Manuscript received February 1, 2019; accepted March 15, 2019. Date of publication April 23, 2019; date of current version October 30, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61471024, in part by the National Marine Technology Program for Public Welfare under Grant 201505002-1, and in part by the National Key Research and Development Program of China under Grant 2017YFC1405600. *(Corresponding author: Haitao Lang.)*

The authors are with the Department of Physics and Electronics, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: haitaolang@hotmail.com; langht@mail.buct.edu.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2907139

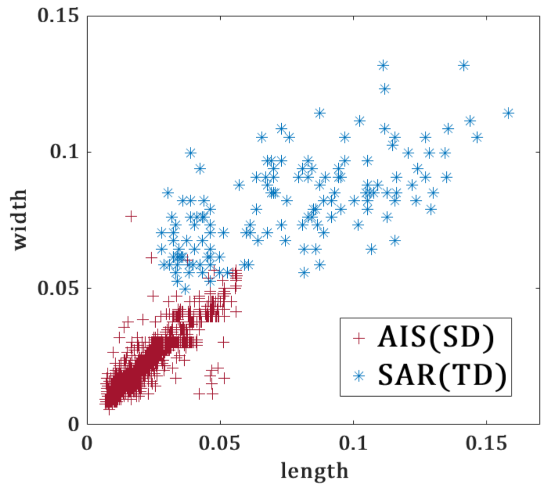

Fig. 1. Distribution difference between AIS and SAR data (ship length and width).

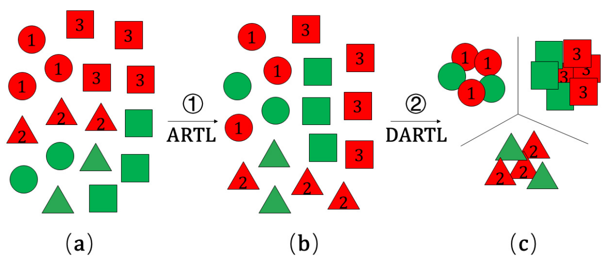

Fig. 2. D-ARTL: an improvement to ARTL. Red: labeled data in SD. Green: unlabeled data in TD. Circle, Square, and Triangle represent three different ship types, respectively. (a) Unlabeled data in TD cannot be classified directly from the labeled data in SD due to mismatched distributions. (b) After ARTL, the domain shift has been greatly reduced, but the discriminative ability of classifier is still weak. (c) After D-ARTL, the intraclass tightness and the interclass separability are improved by using SDIP.

distribution difference and preserve the important properties of the original data. Pan *et al.* [8] proposed the transfer component analysis (TCA) to find a unified transformation that projects data from two domains into a new space to reduce the marginal distribution mismatch. Joint distribution adaptation (JDA) improves TCA by considering not only the marginal distribution shift but also the conditional distribution shift using the pseudo labels of TD [9]. Scatter component analysis (SCA) takes the between- and within-class scatters of SD into consideration [10]. Long *et al.* [11] proposed a general framework, termed as adaptation regularization framework-based transfer learning (ARTL), to model the joint distribution adaptation and manifold regularization in a unified way underpinned by the structural risk minimization (SRM) principle and regularization theory.

In this letter, we propose the method of discriminative ARTL (D-ARTL), which is an improvement to ARTL by introducing a new regularization term of source discriminative information preservation (SDIP). As shown in Fig. 2, ARTL can learn an adaptive classifier by simultaneously optimizing the structural risk function, matching both marginal and conditional distributions, and the manifold consistency, whereas the proposed D-ARTL can further improve the discriminative ability of learned classifier by transferring the SDIP from SD to TD based on the distribution alignment.

II. Discriminative Adaptation Regularization Framework-Based Transfer Learning

A. Problem Definition

A domain consists of a d-dimensional feature space and a marginal distribution, and its corresponding task consists of a label space and a conditional distribution. Given a labeled SD $D_s = \{(\boldsymbol{x}_1^s, \boldsymbol{y}_1^s), \ldots, (\boldsymbol{x}_{n_s}^s, \boldsymbol{y}_{n_s}^s)\}$ and an unlabeled TD $D_t = \{(\boldsymbol{x}_1^t), \ldots, (\boldsymbol{x}_{n_t}^t)\}$. This letter assumes that the feature space and the label space of the SD and the TD are same, but the marginal distribution ($P_s(\boldsymbol{x}^s) \neq P_t(\boldsymbol{x}^t)$) and the conditional distribution ($Q_s(\boldsymbol{y}^s|\boldsymbol{x}^s) \neq Q_t(\boldsymbol{y}^t|\boldsymbol{x}^t)$) are different.

B. ARTL Algorithm

Suppose the classifier f can be represented as

$$f(x) = \boldsymbol{w}^T \phi(\boldsymbol{x}) = \sum_{i=1}^{n_s+n_t} \boldsymbol{\alpha}_i \boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}) \tag{1}$$

where $\boldsymbol{w}$ is the classifier parameters, and $\phi(\boldsymbol{x})$ is the mapping function that projects the original features to a reproducing kernel Hilbert space, n_s and n_t are the number of training data in SD and TD, respectively. $\boldsymbol{\alpha}_i$ is a coefficients and $\boldsymbol{K}$ is a kernel function. The learning framework of ARTL [11] is formulated as

$$f = \arg\min \sum_{i=1}^{n_s} \ell(f(\boldsymbol{x}_i, \boldsymbol{y}_i)) + \sigma \|f\|_{\boldsymbol{K}}^2 + \lambda R_d(D_s, D_t) + \gamma R_m(D_s, D_t) \tag{2}$$

where σ, λ, and γ are the positive regularization parameters. $R_d(\cdot,\cdot)$ denotes the joint distribution adaption between both marginal and conditional distributions, and $R_m(\cdot,\cdot)$ denotes the manifold consistency alignment (MCA). ARTL aims to optimize three complementary objective functions.

1) Minimizing the Structural Risk Function: The ultimate goal of the algorithm is to learn an adaptive classifier for the TD. ARTL adapts the SRM principle and minimize the structural risk function as

$$\begin{aligned}\sum_{i=1}^{n_s} (\boldsymbol{y}_i - f(\boldsymbol{x}_i))^2 + \sigma \|f\|_{\boldsymbol{K}}^2 &= \sum_{i=1}^{n_s+n_t} \boldsymbol{E}_{ii}(\boldsymbol{y}_i - f(\boldsymbol{x}_i))^2 + \sigma \|f\|_{\boldsymbol{K}}^2 \\ &= \|(\boldsymbol{Y} - \boldsymbol{\alpha}^T \boldsymbol{K})\boldsymbol{E}\|_F^2 + \sigma\, tr(\boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha})\end{aligned} \tag{3}$$

where $\boldsymbol{E}$ is a diagonal label indicator matrix with $\boldsymbol{E}_{ii} = 1$ if $\boldsymbol{x}_i \in D_s$, and $\boldsymbol{E}_{ii} = 0$ otherwise. $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_s+n_t}]$ is the label matrix. $tr(\cdot,\cdot)$ is the trace operation.

2) Minimizing the Joint Distribution Difference: Since the SRM principle requires the training and test data to be sampled from identical probability distribution, thus it needs to minimize the difference between the joint probability distributions, i.e., marginal distribution and conditional distribution.

In the implementation, the nonparametric maximum mean discrepancy (MMD) is used as the distribution distance measure

$$R_d(D_s, D_t) = R_{\mathrm{md}}(D_s, D_t) + R_{\mathrm{cd}}(D_s, D_t) \tag{4}$$

where $R_{\mathrm{md}}(\cdot,\cdot)$ and $R_{\mathrm{cd}}(\cdot,\cdot)$ denotes the marginal distribution alignment (MDA) and conditional distribution alignment (CDA), respectively,

$$\begin{aligned} R_{\mathrm{md}}(D_s, D_t) &= \left\| \frac{1}{n_s}\sum_{i=1}^{n_s} f(\boldsymbol{x_i}) - \frac{1}{n_t}\sum_{j=1}^{n_t} f(\boldsymbol{x_j}) \right\|^2 \\ &= tr(\boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{L}_{\mathrm{md}} \boldsymbol{K}\boldsymbol{\alpha}) \end{aligned} \tag{5}$$

$$\begin{aligned} R_{\mathrm{cd}}(D_s, D_t) &= \sum_{c=1}^{C} \left\| \frac{1}{n_s^{(c)}} \sum_{\boldsymbol{x_i}\in D_s^{(c)}} f(\boldsymbol{x_i}) - \frac{1}{n_t^{(c)}} \sum_{\boldsymbol{x_j}\in D_t^{(c)}} f(\boldsymbol{x_j}) \right\|^2 \\ &= \sum_{c=1}^{C} tr(\boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{L}_{\mathrm{cd}} \boldsymbol{K}\boldsymbol{\alpha}) \end{aligned} \tag{6}$$

where

$$L_{\mathrm{md}}(i,j) = \begin{cases} \dfrac{1}{n_s n_s}, & \boldsymbol{x_i},\ \boldsymbol{x_j} \in D_s \\ \dfrac{1}{n_t n_t}, & \boldsymbol{x_i},\ \boldsymbol{x_j} \in D_t \\ \dfrac{-1}{n_s n_t}, & \text{otherwise} \end{cases} \tag{7}$$

$$L_{\mathrm{cd}}(i,j) = \begin{cases} \dfrac{1}{n_s^{(c)} n_s^{(c)}} & \text{if } \boldsymbol{x_i},\ \boldsymbol{x_j} \in D_s^{(c)} \\ \dfrac{1}{n_t^{(c)} n_t^{(c)}} & \text{if } \boldsymbol{x_i},\ \boldsymbol{x_j} \in D_t^{(c)} \\ \dfrac{-1}{n_s^{(c)} n_t^{(c)}}, & \begin{cases} \boldsymbol{x_i} \in D_s^{(c)}, & \boldsymbol{x_j} \in D_t^{(c)} \\ \boldsymbol{x_i} \in D_t^{(c)}, & \boldsymbol{x_j} \in D_t^{(s)} \end{cases} \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

3) Maximizing the Manifold Consistency: To further utilize the information provided by the unlabeled data in TD, ARTL adopts the MCA underlying the marginal distribution

$$\begin{aligned} R_m(D_s, D_t) &= \frac{1}{2}\sum_{i,j=1}^{n_s+n_t} (f(\boldsymbol{x_i}) - f(\boldsymbol{x_j}))^2 \boldsymbol{W_{ij}} \\ &= \sum_{i,j=1}^{n_s+n_t} f(\boldsymbol{x_i}) \boldsymbol{L_m} f(\boldsymbol{x_j}) \\ &= tr(\boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{L_m} \boldsymbol{K}\boldsymbol{\alpha}) \end{aligned} \tag{9}$$

where $\boldsymbol{W_{ij}}$ is the graph affinity matrix between $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$, $\boldsymbol{L_m} = \boldsymbol{I} - \boldsymbol{M}^{-1/2}\boldsymbol{W}\boldsymbol{M}^{-1/2}$ is the graph Laplacian matrix, where $\boldsymbol{M}$ is a diagonal matrix with $\boldsymbol{M_{ii}} = \sum_{j=1}^{n_s+n_t} \boldsymbol{W_{ij}}$, $\boldsymbol{W_{ij}}$ is defined as

$$\boldsymbol{W_{ij}} = \begin{cases} \cos(\boldsymbol{x_i}, \boldsymbol{x_j}), & \boldsymbol{x_i} \in N(\boldsymbol{x_j}) \text{ or } \boldsymbol{x_j} \in N(\boldsymbol{x_i}) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $N(\boldsymbol{x_j})$ denotes the k-nearest neighbor (NN) list of the data point $\boldsymbol{x_j}$.

C. D-ARTL Algorithm

In this letter, we propose to add a novel regularization term $R_s(\cdot,\cdot)$ (SDIP) to the general framework of ARTL. This regularization term preserves the discriminative information of the labeled data in SD. Considering that the joint distribution of both domains are aligned, thus the SDIP should improve further the discriminative ability of the learned classifier. The improved objective function f can be represented as

$$\begin{aligned} f = \arg\min & \sum_{i=1}^{n_s} (\boldsymbol{y_i} - f(\boldsymbol{x_i}))^2 + \sigma \| f \|_{\boldsymbol{K}}^2 \\ & + \lambda R_d(D_s, D_t) + \gamma R_m(D_s, D_t) + \mu R_s(D_s) \end{aligned} \tag{11}$$

where μ is the corresponding regularization parameter.

1) Source Discriminative Information Preservation: Since the labels in the SD are available, we can employ the label information to constrain the new representation of SD data to be discriminative. In this letter, we make the source data share a label (i.e., same class) similar by minimizing the intraclass variance

$$R_s(D_s) = \min\ tr(\boldsymbol{\alpha}^T \boldsymbol{D^s} \boldsymbol{\alpha}) \tag{12}$$

where $\boldsymbol{D^s}$ is the within-class scatter matrix of SD

$$\boldsymbol{D^s} = \sum_{c=1}^{C} \left(\boldsymbol{K_c^s}\right)^T \boldsymbol{H_c^s} \boldsymbol{K_c^s} \tag{13}$$

with the centering matrix $\boldsymbol{H_c^s} = \boldsymbol{I_{n_c^s}} - (1/n_c^s)\boldsymbol{1_{n_c^s}}\boldsymbol{1}_{\boldsymbol{n_c^s}}^T$, where $\boldsymbol{I_{n_c^s}}$ denotes a $n_c^s \times n_c^s$ identity matrix and $\boldsymbol{1_{n_c^s}}$ denotes a vectors of ones. C and c represent the total number of types and one certain type, respectively. Summarizing above, we get

$$R_s(D_s) = tr(\boldsymbol{\alpha}^T \boldsymbol{E}\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}\boldsymbol{E}\boldsymbol{\alpha}) \tag{14}$$

where

$$H = \begin{bmatrix} \boldsymbol{H^s} & 0 \\ 0 & 0 \end{bmatrix} \quad \boldsymbol{H^s} = \begin{bmatrix} \boldsymbol{H_1^s} & 0 & 0 \\ 0 & \boldsymbol{H_c^s} & 0 \\ 0 & 0 & \boldsymbol{H_C^s} \end{bmatrix}.$$

2) Overall Objective Function of D-ARTL:

$$\begin{aligned} f = \arg\min & \|(\boldsymbol{Y} - \boldsymbol{\alpha}^T\boldsymbol{K})\boldsymbol{E}\|_F^2 + tr(\sigma\boldsymbol{\alpha}^T\boldsymbol{K}\boldsymbol{\alpha} \\ & + \boldsymbol{\alpha}^T\boldsymbol{K}(\lambda\boldsymbol{L_d} + \gamma\boldsymbol{L_m} + \mu\boldsymbol{E}\boldsymbol{H}\boldsymbol{E})\boldsymbol{K}\boldsymbol{\alpha}). \end{aligned} \tag{15}$$

where $L_d = L_{md} + L_{cd}$. Setting derivative of objective function as 0, we can obtain

$$\boldsymbol{\alpha} = ((\boldsymbol{E} + \lambda\boldsymbol{L_d} + \gamma\boldsymbol{L_m} + \mu\boldsymbol{E}\boldsymbol{H}\boldsymbol{E})\boldsymbol{K} + \sigma\boldsymbol{I})^{-1}\boldsymbol{E}\boldsymbol{Y}^T. \tag{16}$$

The pseudocode of D-ARTL is given in Algorithm 1.

III. Experiments and Discussion

A. Comparison Methods

We compared the performance of D-ARTL with several state-of-the-art unsupervised DA methods that have been proved on many classification/recognition tasks as shown in Table I, including data-centric methods: TCA [8], JDA [9], balanced distribution adaptation (BDA) [12], visual domain adaptation (VDA) [13], discriminative label consistent (DLC) [14], manifold embedded distribution alignment (MEDA) [15], ARTL [11], and subspace centric methods:

Algorithm 1 D-ARTL Algorithm

Require: Input data $X = [X^s, X^t]$, label Y^s, the hyper-parameters σ, γ, λ, μ and neighbor k;

Ensure: Classifier f

1: Training a classifier with SD data to obtain pseudo-labels of TD data.
2: Choosing a suitable kernel function to build the kernel matrix and map the original space to a high-dimensional space.
3: Constructing MCA matrix according to the formula (9); Constructing SD discriminative information preservation matrix according to formula (14).
4: **repeat**
5: Constructing joint distribution alignment matrix according to formula (4), and normalize $L_d = \frac{L_d}{\|L_d\|_F}$.
6: Calculating $\boldsymbol{\alpha}$ according to formula (16).
7: **until** Convergence
8: Return classifier f by Equation (1).

SCA [10], linear-discriminant-analysis-inspired domain adaptation (LDADA) [16]. TCA performs MDA. JDA goes one step further and proposes to simultaneously align marginal and conditional distributions with equal weights. In addition to adapt both marginal and conditional distributions, BDA also leverages the importance of those two distributions, VDA constructs condensed domain-invariant clusters in the embedding representation to separate various classes, and DLC introduces a repulsive force term to increase interclass distance and, thereby, facilitates discriminative learning. Compared to ARTL, MEDA learns a domain-invariant classifier in Grassmann manifold with SRM, while performing dynamic distribution alignment to quantitatively account for the relative importance of two distributions. SCA converts the data into a set of subspace (i.e., scatters) and then minimizes the divergence between them. LDADA does not model the data distribution to reduce the domain discrepancy but utilize the class mean to learn the class-specific linear projection.

B. Data Preparation and Experimental Protocol

In this letter, we investigated the performance of D-ARTL on the task of ship classification in SAR images (TD) by the assistance of AIS data (SD). The experimental data set followed [7]. For the AIS data, there are 600 labeled samples per class. For the SAR data, there are 50 unlabeled samples per class.[1] Both AIS (SD) and SAR (TD) data share three common ship classes, i.e., cargos, containers, and tankers.

Different from other VDA tasks, since AIS can only provide basic data related to the length and width of the ship, thus the geometric features are only available between domains. Following [7], we extracted NGFs for ship representation. As suggested by the original literature, 1-NN was chosen as the base classifier for TCA, JDA, BDA, SCA, VDA, and DLC. While for LDADA, MEDA, ARTL, and D-ARTL, a domain-invariant classifier was learned according to the corresponding

[1]Since this letter focuses on the unsupervised DA method, we deliberately do not use the labels of TD data.

TABLE I

REGULARIZATION CRITERIA COMPARISON

	TCA	JDA	BDA	SCA	VDA	DLC	MEDA	ARTL	D-ARTL
SRM							✓	✓	✓
MDA	✓	✓	✓	✓	✓	✓	✓	✓	✓
CDA		✓	✓		✓	✓	✓	✓	✓
MCA							✓	✓	✓
DIP				✓	✓	✓			✓

[1] SCA, VDA, DLC and D-ARTL utilize different strategy to preserve the discriminative information of data. For ease of expression, this letter uniformly marks them as discriminative information preservation (DIP).

algorithm. Considering that AIS $\rightarrow$ SAR task is a completely new task for all comparison methods, in order to ensure that each method achieves its best result, the optimal parameters were obtained by a greedy searching in the parameter space.

In order to gain a statistically meaningful understanding of the method on a single DA task like our study, we conducted ten times experiments for each method by varying sampling data in SD. Specifically, for each experiment, we randomly sampled 300 labeled data per class in SD and 50 unlabeled data per class in TD to train the classifier, then tested the learned classifier on TD. The mean and standard deviation of *Accuracy*, which is defined as [11], are used as the evaluation metric

$$Accuracy = \frac{|\{x : x \in D_t \wedge f(x) = y(x)\}|}{|\{x : x \in D_t\}|}. \tag{17}$$

C. Experimental Results and Analysis

The classification accuracy results are given in Table II. From those results, we can make several observations as follows.

First, D-ARTL outperformed all other comparison methods in AIS$\rightarrow$SAR task. The average classification accuracy of D-ARTL was 78.13%. Compared to the best comparison methods ARTL(76.67%) and MEDA(76.00%), the average performance improvement was 1.46% and 2.13%, respectively, corresponding to a significant average error reduction of 6.26% and 8.88%, respectively. This improvement stems from the introduction of the proposed SDIP regularization item to the DA framework. The role of DIP can also be found by comparing JDA(52.67%) with DLC(54.40%), it demonstrates that under the same (or nearly equal) conditions, the introduction of DIP item can further boost the performance.

Second, observing the results of TCA(49.60%), JDA(52.67%), DLC(54.40%), ARTL(76.67%), and D-ARTL(78.13%), it shows that with the increasing regularization terms, the classification performance continues to increase. The method of distribution importance balance can further improve the performance of distribution adaptation, which is shown by comparing JDA(52.67%) with BDA(58.00%). In order to inspect the effectiveness of a regularization term in (11) to the classification accuracy, we run D-ARTL with and without corresponding regularization term (JDA, MCA, and SDIP), and compute the accuracy decline. Table III illustrates the classification accuracy and corresponding decline after regularization term removal. We observe that in AIS $\rightarrow$ SAR task, removing JDA leads to the accuracy decline up to 19.86%. It implies that there exists a large distribution divergence between

TABLE II
MEAN AND STANDARD DEVIATION OF ACCURACY(%)

TCA	JDA	BDA	SCA	VDA	LDADA	DLC	MEDA	ARTL	D-ARTL
49.60±2.81	52.67±3.09	58.00±2.21	43.07±3.45	51.33±6.75	45.47±0.99	54.40±3.52	76.00±2.05	76.67±2.79	78.13±4.20

TABLE III
PERFORMANCE DECLINE AFTER REGULARIZATION TERM REMOVAL(%)

	-JDA	-MCA	-SDIP
acc.± std.	58.27± 4.23	72.93±6.16	76.67±2.79
Decline	-19.86	-5.2	-1.46

AIS and SAR domains as shown in Fig. 1. The removal of MCA brings a decline of 5.2%, it clearly indicates the important role of manifold consistency underlying marginal distribution. The decline of 1.26% does not mean that SDIP has little effect on classifier. Instead, it shows that ARTL can be further improved by introducing SDIP to a higher level. A 5.6% reduction in error is very critical to many real application scenarios.

Third, the results obtained by the two subspace centric methods, SCA (43.07%) and LDADA (45.47%), are lower than our expectation. For SCA, a possible explanation is that although SCA considers the total scatter, domain scatter, and class scatter using a unified mapping, however, there may not exist such a common subspace between AIS and SAR domains, which satisfies all the constrains. As for LDADA, a premise for it to work well is that a good target annotation can be readily acquired with a source-trained model given sufficiently good feature representations. Consider the NGFs used in this letter is not as good as deep features used in visual domain, this may be a possible reason that degenerates the performance of LDADA on AIS $\rightarrow$ SAR task.

D. Other SDIP Method

There are several methods to implement SDIP, for instance, VDA and D-ARTL force the intraclass tightness, DLC enhances the interclass distance. In this section, to inspect the performance of different SDIP strategies, we also implemented another typical method of SDIP by minimizing the intraclass distance and maximizing the interclass distance simultaneously, which is formulated as

$$R_s(D_s) = \min\ tr\left(\boldsymbol{\alpha}^T\left(\boldsymbol{D}_s^s - \boldsymbol{D}_d^s\right)\boldsymbol{\alpha}\right) \tag{18}$$

where $\boldsymbol{D}_s^s$ is the sum of distances from similar class and $\boldsymbol{D}_d^s$ is the sum of distances from dissimilar class. By replacing the corresponding item in (11) and repeating the experiments following the same protocol, we achieved a competitive performance in terms of mean and standard deviation of accuracy 76.80 ± 1.45. This result indicates that the performance of D-ARTL stays robust with different SDIP strategy choices.

IV. CONCLUSION

In order to meet the urgent needs of current marine management for ship monitoring capabilities, this letter aims to improve ship classification performance in SAR image in case that there is no labeled data available. To this end, we propose to improve a general unsupervised DA framework ARTL by adding a novel SDIP regularization term. The experiments demonstrate that D-ARTL outperforms state-of-the-art methods on the task of AIS $\rightarrow$ SAR. To the best of our knowledge, D-ARTL is the first attempt to performance joint optimization to SRM, JDA, MCA, and SDIP.

In the future, we plan to further improve D-ARTL by adding new constrains, such as distribution importance balance. We also attempt to use D-ARTL for more complex feature adaptation, such as depth features, and for a wider range of VDA tasks.

REFERENCES

[1] K. Ji, X. Xing, W. Chen, H. Zou, and J. Chen, "Ship classification in TerraSAR-X SAR images based on classifier combination," in *Proc. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2013, pp. 2589–2592.

[2] X. Xing, K. Ji, H. Zou, W. Chen, and J. Sun, "Ship classification in TerraSAR-X images with feature space based sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1562–1566, Nov. 2013.

[3] H. Lang, J. Zhang, X. Zhang, and J. Meng, "Ship classification in SAR image by joint feature and classifier selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 212–216, Feb. 2016.

[4] H. Lang and S. Wu, "Ship classification in moderate-resolution sar image by naive geometric features-combined multiple kernel learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1765–1769, 2017.

[5] C. Bentes, D. Velotto, and B. Tings, "Ship classification in TerraSAR-X images with convolutional neural networks," *IEEE J. Ocean. Eng.*, vol. 43, no. 1, pp. 258–266, May 2018.

[6] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.

[7] H. Lang, S. Wu, and Y. Xu, "Ship classification in SAR images improved by AIS knowledge transfer," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 439–443, May 2018.

[8] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[9] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2013, pp. 2200–2207.

[10] M. Ghifary, D. Balduzzi, W. B. Kleijn, and M. Zhang, "Scatter component analysis: A unified framework for domain adaptation and domain generalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1414–1430, Jul. 2017.

[11] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation regularization: A general framework for transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1076–1089, May 2014.

[12] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, May 2017, pp. 1129–1134.

[13] J. Tahmoresnezhad and S. Hashemi, "Visual domain adaptation via transfer feature learning," *Knowl. Inf. Syst.*, vol. 50, no. 2, pp. 585–605, Jun. 2017.

[14] L. Luo *et al.* (2018). "Discriminative label consistent domain adaptation." [Online]. Available: https://arxiv.org/abs/1802.08077

[15] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, and P. S. Yu, "Visual domain adaptation with manifold embedded distribution alignment," in *2018 ACM Multimedia Conf. Multimedia Conf.*, May 2018, pp. 402–410.

[16] H. Lu, C. Shen, Z. Cao, Y. Xiao, and A. van den Hengel, "An embarrassingly simple approach to visual domain adaptation," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3403–3417, Jun. 2018.