Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)
Paris, France - August, 2025
Paper No. 105
DOI: 10.11159/mvml25.105

Ship Detection for Satellite Images based on Classifier Transfer Learning Combined with Feature Transfer Learning

Huan Zhang[1], Qianglin Liu[2], Xiaolin Han[3], Lijuan Niu[2], and Weidong Sun[1]
[1] Department of Electronic Engineering, Tsinghua University, Beijing, China
zhanghuan19@tsinghua.org.cn; wdsun@tsinghua.edu.cn
[2] Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
qianglinliu0927@163.com; niulijuan8197@126.com
[3] School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China
hxl@bit.edu.cn

Abstract - Transfer learning (TL) is a powerful tool to transfer deep learning models from a large source dataset to a small target dataset, but the upper-layers of deep learning models are less transferable for lacking universality and possessing specificity to certain tasks. Most researches have focused on feature-oriented transfer learning base on the feature space, however, both the classifier-oriented transfer learning and the label space haven't been considered. Faced with these issues, a generalized classifier-oriented transfer learning, termed as classifier-TL, is proposed in this paper, which investigates the correlation between source label space and target label space to transfer and refine the generalized classifier. More specifically, for a given task, a label space descriptor is proposed to depict the label space, and a label space similarity is introduced to measure the correlation between source label space and target label space. Then, the target label space is focused through the proposed label driven posteriori optimization, trying to exploit similar label spaces of the closest category. In this procedure, the classifier can be refined from a set of generalized classifiers to a specific classifier. Furthermore, this classifier-TL can be combined with the traditional feature-oriented transfer learning, to form an integrative secondary transfer learning, for further boosting the performance of transfer learning. Experimental results for the task of ship detection, have demonstrated the effectiveness of our proposed method.

***Keywords*:** classifier transfer learning, secondary transfer learning, label space, transfer learning, CNN

1. Introduction

Deep learning models have made numerous achievements in computer vision field [1]-[3], due to their excellent feature extraction ability [4], which is particularly prominent in transfer learning (TL). However, the upper-layers of deep learning models are less transferable for lacking generality and possessing specificity to certain tasks [5]. Most researches have focused on feature-oriented transfer learning, *i.e.* transfer the feature extraction ability, ignoring the flip side of transfer learning, *i.e.* the classifier-oriented transfer learning.

While in machine learning methods, the algorithm except feature extraction can be considered as a classifier. Similarly, from the perspective of input and output, in the deep learning models, network structure between the output of feature extractor and the output of the entire model plays the role of a classifier [6], which can be considered and termed as a generalized classifier Ψ_D. And the generalized classifier Ψ_D is not limited to classification task, which only consists of a classification branch, but may also consists of RPN, RoI pooling/RoI Align, convolution layers, activation layers, fully connection layers and other network structures deployed for different tasks. These structures are not designed to extract features, but to map the extracted features from feature space to label space for prediction.

The transfer of feature extraction ability, *i.e.* the traditional feature-oriented transfer learning (Feature-TL), exploits the relation between source feature space and target feature space, utilizing the instance reweighting and reusing, feature mapping and alignment, parameter fine-tuning and relation mapping, *etc* [7]. But few methods have considered the correlation between the source label space and the target label space [8]. In addition, the universality and the transferability of generalized classifier hasn't been paid attention.

Faced with the above issues, a generalized classifier-oriented transfer learning method, termed as classifier-TL, is proposed in this paper. More specifically, for a given task, a label space descriptor is proposed to depict the label space, and a label space similarity is introduced to measure the correlation between the source label space and the target label space. Then, the target label space is focused by the proposed label driven posteriori optimization, trying to use the label space of the closest categories more effectively. In this procedure, the classifier can be refined from a set of generalized classifiers to a specific classifier. Furthermore, this classifier-TL can be combined with traditional feature-TL, to form an integrative secondary transfer learning, termed as secondary-TL, for further boosting the effects of transfer learning. Experiments for the task of ship detection in remotely sensed satellite images, have demonstrated the effectiveness of our proposed method. The novelties and the contributions of our method are as follows:

1) To our best knowledge, this is the first time that the generalized classifier-oriented transfer learning is proposed, in which the universality and the transferability of generalized classifier is concentrated. On this basis, the secondary transfer learning is firstly proposed, which combines the feature-TL and the classifier-TL effectively.

2) This is the first time that the label space descriptor is proposed, based on which a label space similarity is introduced. Then, the focus of label spaces is firstly proposed, exploiting the proposed label driven posteriori optimization.

2. Related Works

According to the question, whether the labels of target domain are accessible, transfer learning can be categorized into three cases [9]. The first case is inductive transfer learning, in which the labels of target domain are totally accessible. The second case is the transductive transfer learning, in which the labels of target domain is partially accessible. And, the third case is the unsupervised transfer learning, in which the labels of target domain are totally inaccessible. An implicit assumption of the second and the third case is that, target categories in source domain and that in the target domain are the same, which is also a limitation for these methods.

Furthermore, transfer learning itself concerned with all the above cases can basically be summarized into one type, *i.e.* the feature-oriented transfer learning, as for the universality and the transferability of generalized classifier, we can only find several implicit efforts towards this direction. Among them, PathNet utilized embedded agents, to discover which part of the network can be re-used for new tasks [11]. DAN embedded task-specific layers in a reproducing kernel Hilbert space, where the mean embedding of different domain distributions can be matched [12]. RTN utilized a domain adaptation approach, which learned adaptive classifiers and transferable features from labelled data in source domain, to unlabelled data in the target domain [13]. DTA leveraged adversarial dropout to learn discriminative features by enforcing the cluster assumption and designing objective functions, to support robust domain adaptation [14]. [15] put forward a sequential Monte Carlo filter, to select relevant samples and estimate unknown target distribution for learning a classifier. Domain Adaptive Faster R-CNN improved cross-domain robustness of detector on both image level and instance level [16]. However, none of them considered the label space in transfer learning, the universality and transferability of generalized classifier, or the relation between them.

3. Method

3.1. Background

The notation of transfer learning should be given at first. Generally, the domain of the training data and that of the test data can be denoted as source domain $\boldsymbol{D}^S$ and $\boldsymbol{D}^T$, respectively. More generally, A domain $\boldsymbol{D}$ can be defined by the feature space $\boldsymbol{X}$ and the marginal probability distribution $P(x)$, where $x_i = \{x_i^1, x_i^2, ..., x_i^n\} \in \boldsymbol{X}$, x_i denotes a training sample, and x_i^n denotes all possible features. Given a domain $\boldsymbol{D} = \{\boldsymbol{X}, P(x)\}$, a learning task can be denoted as $\boldsymbol{L} = \{\boldsymbol{Y}, f(x)\}$, where $\boldsymbol{Y}$ denotes the ground truth label space, $\boldsymbol{Y} = \{y_1, y_2, ..., y_n\}$, $f(x)$ denotes the prediction function that can be learned from labelled pairs $\{x_i, y_i\}$, $x_i \in \boldsymbol{X}$ and $y_i \in \boldsymbol{Y}$ [13], can be written as a conditional probability distribution $P(y|x)$. So, source domain $\boldsymbol{D}^S$ and target domain $\boldsymbol{D}^T$ can be further described as $\boldsymbol{D}^S = \{\boldsymbol{X}^S, P(x)^S\}$ and $\boldsymbol{D}^T = \{\boldsymbol{X}^T, P(x)^T\}$, respectively. It is acknowledged that, given a task $\boldsymbol{L}^T$ with $\boldsymbol{D}^T$, the prediction function $f(x)$ can be boosted by transferring the learned knowledge from the $\boldsymbol{L}^S$ with $\boldsymbol{D}^S$ [4].

As discussed in related works [8], this definition of transfer learning is inadequate. Theoretically, from the perspective of Bayes formulation, $P(x)P(y|x) = P(y)P(x|y)$, we can see that, the left part of this equation corresponding to the traditional feature-TL, while for the right part of this equation, *i.e.* the marginal probability distribution of label space $P(y)$ and the significance of the conditional probability distribution $P(x|y)$ has never been deeply considered.

3.2. Problem Definition

To illustrate the marginal probability distribution of label space $P(y)$ and the conditional probability distribution $P(x|y)$ more clearly, here, the task of object detection in remotely sensed images is taken as an example. Similar with the prediction function $f(x)$ that can be learned from the conditional probability distribution $P(y|x)$, in the deep learning based object detection task, it can generally be considered as the learning of a posterior $P(C,B|I)$ [16], where $C=\{C_1,C_2,...,C_K\}$ denotes categories of the targets, B denotes the bounding-boxes and I denotes the image representation, so the joint distribution of training samples for object detection can be denoted as $P(C,B,I)$. Due to the domain divergence between source domain and target domain, *i.e.* $\boldsymbol{D}^S \neq \boldsymbol{D}^T$, the joint distribution of $\boldsymbol{D}^S$ and $\boldsymbol{D}^T$ are unequal, *i.e.* $P(C^S,B,I) \neq P(C^T,B,I)$. In transfer learning based object detection, the domain divergence between $\boldsymbol{D}^S$ and $\boldsymbol{D}^T$ should be as minimum as possible, which can be formulated as:

$$\Gamma = min||P(C^S,B,I) - P(C^T,B,I)|| \tag{1}$$

where, the joint distribution can be decomposed by Bayes formulation:

$$P(C,B,I) = P(C|B,I)P(B,I) \tag{2}$$

Combined with Eq.(1), we can get:

$$\Gamma = min||P(C^S|B,I)P(B,I) - P(C^T|B,I)P(B,I)|| \tag{3}$$

If the $P(B,I)$ is further decoupled into $P(B,I) = P(B|I)P(I)$, both the marginal probability distributions $P(I)$, *i.e.* the distribution of feature representation, and the conditional probability distribution $P(B|I)$, *i.e.* the bounding box prediction, are biased. Therefore, to acquire the domain independent bounding box prediction, Eq.(3) can be converted into the Eq.(4):

$$\Gamma = min||P(B|C^S,I)P(C^S,I) - P(B|C^T,I)P(C^T,I)|| \tag{4}$$

In which, both $P(B|C^S,I)$ and $P(B|C^T,I)$ are domain-independent bounding box prediction. Therefore, by enforcing the consistency between $P(C^S,I)$ and $P(C^T,I)$, the domain divergence between $\boldsymbol{D}^S$ and $\boldsymbol{D}^T$ can be eliminated.

Similarly, $P(C,I)$ can be further decomposed into $P(C,I) = P(C|I)P(I)$, the marginal probability distributions $P(I)$ corresponds to $P(x)$. And in fact, the conditional probability distribution $P(C|I)$ represents mapping from the feature space to the label space, which corresponds to the $P(y|x)$. Both of the above can be constrained by certain similarity metrics, such as MMD, to eliminate the domain divergence. However, in the decomposition, both $P(y)$ and $P(x|y)$ haven't been considered.

Another possible decomposition of $P(C,I)$ can be expressed as $P(C,I) = P(I|C)P(C)$. In this way of decomposition, $P(I|C)$ represents the domain invariant feature expression, which corresponds to the conditional probability distribution $P(x|y)$, and $P(C)$ represents the label space, which corresponds to the marginal probability distribution $P(y)$. The significance and the impacts of these distributions will be detailed, in the following section.

3.3. Generalized Classifier-oriented Transfer Learning

Here, to depict the marginal probability distribution of label space $P(y)$, a label space descriptor will be proposed firstly, and the task of object detection in remote sensed images is taken as an example. In the remotely sensed object detection, we assume that similar sensors, similar data acquisition approaches and similar object patterns could make the remotely sensed target of different categories possess certain similarity in complexity, distribution, appearance and scale [18], which corresponds to the inherent properties of each category and can be depicted by the label space descriptor. In detail, the label space descriptor can be formulated by a combination index Υ_k of different impact factors for the category k , including the averaged target size ρ, averaged target aspect ratio v, target number n and image number N:

$$\Upsilon_k = \frac{\sqrt{\rho_k}}{v_k - \frac{1}{1/C\sum v_k}} \times log\left(\frac{n_k}{N_k}\right), k = 1,...,C_K \tag{5}$$

where, k denotes categories in the dataset. All these parameters are calculated from the annotated labels directly. And, for the reason why only annotated labels are used to evaluate the combination index is that, from the perspective of causality, only the annotated information can be used before the object detection and final object recognition. Taking the most popular remote sensing object detection dataset, the DOTA dataset [20], as an example, Υ_k of each of 15 categories are as shown in Table 1.

Table 1: Combination index Υ_k of different categories in the DOTA dataset.

Category name	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HE
Index value	283.3	143.6	17.0	63.5	30.1	23.7	34.6	64.0	65.2	154.5	100.9	106.4	46.5	47.7	41.0

On this basis, the relationship between the combination index and the object detection accuracy on the DOTA dataset is plotted in Figure 1, in which the x-axis is log of Υ_k and the y-axis is calculated by 12 representative object detection methods, including RT, R3Det, BBA vector, *etc* [23].

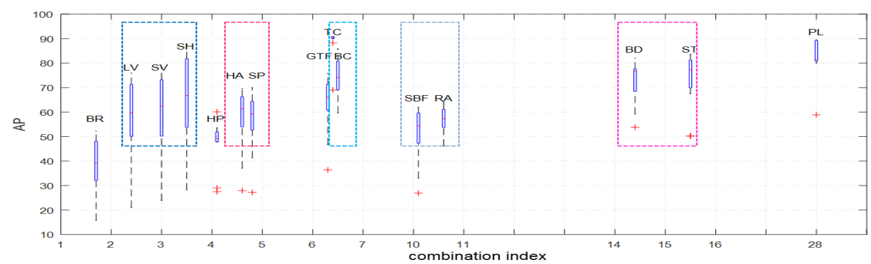

Fig. 1: Relation between combination index and the detection accuracy of different categories.

From Fig.1, it can be seen that, if two category i and j possess certain similarity in complexity, distribution, appearance and scale, their Υ_i and Υ_j are also close. Therefore, a label space similarity $L_{i,j}$ is introduced to measure the similarity between the label space of category i and category j:

$$L_{i,j} = e^{-0.1\|\Upsilon_i - \Upsilon_j\|_1},\ i,j = 1,\dots,K \tag{6}$$

where, $L_{i,j} \in (0,1]$. And, the $L_{i,j}$ becomes large, when the label spaces of different categories possess certain similarity, such as small vehicle (SV) vs ship (SH) and tennis court (TC) vs basketball court (BC).

Similar with the feature-oriented transfer learning, it can be assumed that the label space can be focus by taking the advantage of correlation between the source label space and the target label space. And, the classifier can be transferred and refined from a set of generalized classifiers to a specific classifier, by our proposed label driven posteriori optimization.

More detailed, to drive the posteriori optimization through similarity between the source label space and the target label space, there is a label space focus assumption, *i.e.* the target label space Y^T can be focused from the source label space Y^S. To focus the target label space, the label space of target categories $\{x_i^T, y_i^T\}$, and the selected label space of the most similar categories $\{y_i^{T'}\}$ by the label space descriptor, are utilized to train the network. It worth mentioned that, in the label space, the most similar category is only with the label $\{y_i^{T'}\}$ but the image $\{x_i^{T'}\}$, that could lead to negative transfer effects of that category [19]. Due to the label space similarity of these most similar categories, the label space of target category can be focused, which could lead to the transfer and refine of the classifier Ψ_D from a set of generalized classifiers to a specific classifier. This procedure can be termed as the label driven posteriori optimization, and can be formulated as a maximum a posterior (MAP) procedure given $\{x_i^T, y_i^T\}$ and $\{y_i^{T'}\}$:

$$\theta = \arg\max p\{\theta_D | x_i^T, y_i^T, y_i^{T'}\} \tag{7}$$

where, $p\{\theta_D | x_i^T, y_i^T, y_i^{T'}\}$ is the posterior of the generalized classifier parameter θ_D given training data $x_i^T, y_i^T, y_i^{T'}$. This formulation can be rewritten as:

$$\begin{aligned} p\{\theta_D | x_i^T, y_i^T, y_i^{T'}\} &\propto p\{\theta_D, x_i^T, y_i^T, y_i^{T'}\} \\ &\propto p\{y_i^T | \theta_D, x_i^T, y_i^{T'}\} p\{\theta_D | x_i^T, y_i^{T'}\} \end{aligned} \tag{8}$$

The optimization in Eq.(8) is equivalent to:

$$\hat{\theta} = argmax\, p\{y_i^T/\theta_D, x_i^T\}p\{\theta_D/x_i^T, y_i^{T'}\} \tag{9}$$

In which, $p\{y_i^T/\theta_D, x_i^T\}$ denotes the conventional supervised learning task, $p\{\theta_D/x_i^T, y_i^{T'}\}$ represents the influence of the label space of the most similar categories, which would lead to the label driven generalized classifier-oriented transfer learning. Through this procedure, the classifier of specific task inherits the prediction borders of other categories classifiers, *i.e.* generalized classifiers, and further improves the prediction ability of specific target, here, we called it as classifier-TL.

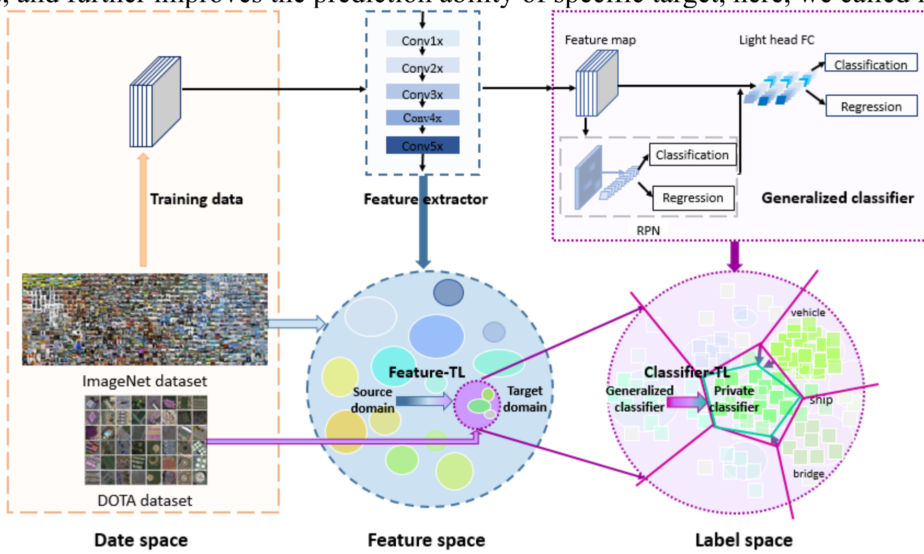

Fig. 2: The relationship between data space, feature space and the label space, and that between feature-TL and the classifier-TL.

3.4. Secondary Transfer Learning

Furthermore, as shown in Fig.2, the above classifier-TL can be combined with the traditional feature-TL, forming an integrative secondary transfer learning, termed as secondary-TL, to further boost the effects of transfer learning. More detailed, the label space focus assumption of secondary transfer learning is similar to that of the generalized classifier-oriented transfer learning. Training data for feature-oriented transfer learning and generalized classifier-oriented transfer learning are denoted as $\{x_i^{T2}, y_i^{T2}, y_i^{T2'}\}$. When conducting the secondary transfer learning, feature-TL and the classifier-TL are trained on $\{x_i^{T2}, y_i^{T2}, y_i^{T2'}\}$ simultaneously, *i.e.* the proposed secondary transfer learning consists of both the feature extractor Φ_E and the generalized classifier Ψ_D. The learnable parameters of the proposed secondary transfer learning are $\theta = \{\theta_E, \theta_D\}$, including the parameters of the feature extractor θ_E and that of the generalized classifier θ_D. This procedure can be taken as the label driven posteriori optimization, and can be formulated as a MAP procedure as well, given the training data $\{x_i^{T2}, y_i^{T2}, y_i^{T2'}\}$:

$$\hat{\theta} = argmax\, p\{\theta/x_i^{T2}, y_i^{T2}, y_i^{T2'}\} \tag{10}$$

where, $p\{\theta/x_i^{T2}, y_i^{T2}, y_i^{T2'}\}$ denotes the posterior of model parameter θ given training data $x_i^{T2}, y_i^{T2}, y_i^{T2'}$. This formulation can be rewritten as:

$$\begin{aligned} p\{\theta/x_i^{T2}, y_i^{T2}, y_i^{T2'}\} &\propto p\{\theta, x_i^{T2}, y_i^{T2}, y_i^{T2'}\} \\ &\propto p\{y_i^{T2}/x_i^{T2}, y_i^{T2'}, \theta\}p\{\theta/x_i^{T2}, y_i^{T2'}\} \end{aligned} \tag{11}$$

The optimization in Eq.(11) is equivalent to:

$$\hat{\theta} = argmax\, p\{y_i^{T2}/x_i^{T2}, \theta\}p\{\theta/x_i^{T2}, y_i^{T2'}\} \tag{12}$$

Similarly, $p\{y_i^{T2}/x_i^{T2}, \theta\}$ denotes the conventional supervised learning. From an optimization perspective, the term $p\{\theta/x_i^{T2}, y_i^{T2'}\}$ plays more significant role in secondary transfer learning, which can be further decomposed:

$$p\{\theta|x_i^{T2},y_i^{T2'}\} = p\{\theta_E\theta_D|x_i^{T2},y_i^{T2'}\}$$
$$\propto p\{\theta_E\theta_D x_i^{T2},y_i^{T2'}\}$$
$$\propto p\{\theta_E|x_i^{T2},y_i^{T2'}\}p\{\theta_D|x_i^{T2},y_i^{T2'}\}$$

In which, $p\{\theta_E|x_i^{T2},y_i^{T2'}\}p\{\theta_D|x_i^{T2},y_i^{T2'}\}$ represents the secondary transfer learning, which depicts the influence of the label space of the most similar category on the both the generalized classifier and the feature extractor.

4. Experiment

4.1. Experiment Setting

To evaluate the effectiveness of our proposed method, experiments are conducted on the remote sensing object detection dataset, the DOTA dataset [20], *i.e.* one of the largest remote sensing object detection datasets, which contains 2806 images of 15 categories collected from different remote sensing platforms, including 188,282 instances. All the instances are of various scales, orientations and shapes.

The Light-Head Faster R-CNN OBB (LR-O) [21], the Deformable Convolutional Networks (DCN) [22] and the RoI Transformer (RT) [23] are selected as the baseline detection method to validate the proposed methods. Among them, LR-O is widely applied to different tasks on object detection in remote sensed images, DCN has broad impact on object detection in the natural images, and RT is the official object detection method for the DOTA dataset.

4.2. Implement Details

In our proposed method, the networks are pre-trained on the ImageNet dataset to achieve the general feature extraction ability. Then, the pre-trained networks will be transferred to the remote sensing dataset, to narrow the gap between natural images and remotely sensed images. Finally, the fine-tuned networks will be transferred to the ship dataset, to accomplish the specific target detection task. And, the ship dataset of specific target task is created and augmented from the DOTA dataset by using the BIRD data augmentation method [1].

The backbone of the object detection networks is the ResNet 101 [24]. After feature extraction, all feature maps will be feed into the RPN, where the horizontal anchors are utilized to proposal RoI (region of interest). Considering the trade-off between the time-consuming and precision, the scale and ratio of the anchors are set to $\{4^2,8^2,16^2,32^2,64^2\}$ and $\{1/2, 1/1, 2/1\}$ respectively. Due to the variant scale, arbitrary orientation and diverse shape of the remote sensing targets, centre coordinates, width, height and angle (x,y,w,h,θ) [25] are utilized to depict the bounding box of target.

In training, a mini-batch is set to 100, the learning rate is set to 0.0005 with the moment set to 0.9, the SGD is utilized as the optimizer, and the mean average precision (mAP) is used as the evaluation criterion.

4.3. Label Space Similarity

In this part, the ablation studies of the label space similarity will be conducted firstly. The ship category is selected as the representative remote sensing target to detect. LR-O, DCN and RT are selected as the baseline detection method. As the label space similarity $L_{i,j}$ is utilized to evaluate the correlation between the label spaces of different categories, the similarities between the label space of ship category and that of the other categories are as shown in Table 2.

Table 2: Label space similarities between ship and other categories in DOTA dataset.

	PL	BD	BR	GTF	SV	LV	TC	BC	ST	SBF	RA	HA	SP	HE
SH														

From Table 2, it can be straightforward seen that, small vehicle (SV) category is most similar to the ship category. Therefore, the labelled ship images and small-vehicle labels without images are combined together, termed as SH+SV. Apart from these, large vehicle (LV) category is medium most similar to the ship category. Therefore, the labelled ship images, small-vehicle labels and large-vehicle labels without images can be combined, termed as SH+SV+LV. In addition, bridge (BR) category is small most similar to the ship category, therefore, the labelled ship images, small-vehicle labels without images, large-vehicle labels without images, and bridge labels without images are also combined, termed as

SH+SV+LV+BR. The detection results are shown in Table 3. It can be seen that, the most similar category contributes most to the results. Therefore, only most similar category is selected for secondary-TL.

Table 3: The influence of similar category number on the performance of secondary transfer learning.

	Shiponly	SH+SV	SH+SV+LV	SH+SV+LV+BR
RT	74.15	**75.94**	75.79	75.40
DCN	63.45	63.70	**63.93**	63.73
LR-O	52.53	**53.17**	52.26	52.18

4.4. Classifier-TL and Secondary-TL

To evaluate the effectiveness of classifier-TL and secondary-TL, the aforementioned methods are compared on the specific ship detection task on the DOTA dataset. In the classifier-TL, the entire feature extractor and the generalized classifier is fine-tuned on ship target dataset with the most similar category labels. On this basis, to validate the effectiveness of the classifier-oriented transfer learning, the upper layers of feature extractor are fine-tuned along with the generalized classifier on the ship target dataset as well. The accuracy of ship detection results with LR-O, DCN and RT methods are shown in Table 4.

The first experiment is conducted with the LR-O method. As in Table 4, the accuracy of directly transfer from ImageNet to the ship category is 43.29%, suffering from data insufficiency compared with that of the primary transfer learning, whose accuracy is 50.50%. And the accuracy of feature-TL twice is 49.54%, which is lower than the primary transfer learning, and can be explained by the negative transfer effects mentioned above. The accuracy of classifier-TL is 1.5% higher than the primary transfer learning, and 2.5% higher than the feature-TL twice. In addition, the accuracy of LR-O fine-tuned by secondary-TL is 3% higher than the primary transfer learning, 4% higher than the feature-TL twice, and 1% higher than classifier-TL.

The second experiment is conducted with the DCN method. As shown in Table 4, the accuracy of feature-TL twice is lower than that of the primary transfer learning, and classifier-TL outperforms feature-TL twice by 3.6%. In addition, the accuracy of DCN fine-tuned by secondary-TL outperforms the directly transfer learning, primary transfer learning and feature-TL twice by 7.9%, 0.5% and 4.4% respectively, which has verified the effectiveness of secondary-TL.

The third experiment is taken with the RT method. As shown in Table 4, the accuracy of the classifier-TL is 1.2% higher than that of the directly transfer learning, 0.4% higher than that of the primary transfer learning, and 0.2% higher than that of the feature-TL twice. The accuracy of RT fine-tuned by secondary-TL is 1.8% higher than that of the primary transfer learning, and 1.5% higher than that of the feature-TL twice. In addition, the accuracy of secondary-TL also outperforms the classifier-TL by 0.8%.

From Table 4, it can be drawn that the classifier-TL and the secondary-TL can take effect across different deep learning models, and outperform both directly training, primary transfer learning and feature-TL twice remarkably, which have demonstrated the effectiveness of both the classifier-TL and the secondary-TL.

Table 4: Comparison results between feature-TL, classifier-TL and secondary-TL with LR-O, DCN and RT methods.

Method	Type of transfer	Description	Accuracy
LR-O	Feature transfer learning	ImageNet->Ship	43.29%
		ImageNet->DOTA (Ship in DOTA)	50.50%
	Feature transfer learning twice	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	49.54%
	Classifier transfer learning	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	52.03%
	Secondary transfer learning	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	**53.17**%
DCN-ResNet101	Feature transfer learning	ImageNet->Ship	56.06%
		ImageNet->DOTA (Ship in DOTA)	63.45%
	Feature transfer learning twice	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	59.56%
	Classifier transfer learning	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	63.15%
	Secondary transfer		**63.93**%

	learning	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	
RT-ResNet101	Feature transfer learning	ImageNet->Ship	73.34%
		ImageNet->DOTA (Ship in DOTA)	74.15%
	Feature transfer learning twice	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	74.39%
	Classifier transfer learning	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	74.58%
	Secondary transfer learning	Step 1: ImageNet->DOTA Step 2: DOTA->Ship	**75.94%**

5. Conclusion

Faced with the ignorance of generalized classifier and label space in the traditional transfer learning, a generalized classifier-oriented transfer learning is proposed in this paper, which exploits the correlation between source label space and target label space, to transfer and refine the generalized classifier. More specifically, for a given task, a label space descriptor is proposed to depict the label space, and a label space similarity is introduced to measure the correlation between source label space and target label space. Then, the target label space is focused by the proposed label driven posteriori optimization, to exploit the most similar source label space. On the basis, an integrative secondary transfer learning is proposed, which combines the feature-TL and classifier-TL, to further boost the performance of transfer learning. To our best knowledge, this is the first time that the classifier-TL and the secondary-TL are proposed, and the label space is firstly depicted and focused. Experimental results for the task of ship detection, have demonstrated the effectiveness of our proposed classifier-TL and the secondary-TL method.

Acknowledgements

This work was supported in part by the National Natural Science Foundation (41971294, 82471999), the Beijing Natural Science Foundation (3254044) and the Beijing Institute of Technology Research Fund Program for Young Scholars of China.

References

[1] H. Zhang, Z. Xu, X. Han, and W. Sun. “Data Augmentation using Bitplane Information Recombination Model”, *IEEE Transactions on Image Processing*, vol.31, pp.3713-3725, 2022.

[2] H. Zhang, X. Han, J. Deng, and W. Sun. “How to Evaluate and Remove the Weakened Bands in Hyperspectral Image Classification”, *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[3] H. Zhang, X. Han, and W. Sun. “Region-Focusing Data Augmentation via Salient Region Activation and Bitplane Recombination for Target Detection”, *Remote Sensing*, vol.16, pp.4806, 2024.

[4] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, and H. Zhu. “A Comprehensive Survey on Transfer Learning”, *Proceedings of the IEEE*, vol.109, pp.43-76, 2019.

[5] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. “How Transferable Are Features in Deep Neural Networks?”, *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[6] H. Zhang, Z. Xu, X. Han, and W. Sun. “Refining FFT-based Heatmap for the Detection of Cluster Distributed Targets in Satellite Images”, *British Machine Vision Conference (BMVC)*, Virtual, 2021.

[7] L. Zhang and X. Gao, “Transfer Adaptation Learning: A Decade Survey”. *IEEE Transactions on Neural Networks and Learning Systems*, vol.35, pp.23-44, 2019.

[8] H. Zhang, W. Leng, X. Han, and W. Sun. “MOON: Subspace-based Multi-branch Network for Object Detection in Remotely Sensed Images”, *Remote Sensing*, vol.15, pp.4201, 2023.

[9] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu. “A survey on deep transfer learning”, *International Conference on Artificial Neural Networks*. Springer, Cham, pp.270-279, 2018.

[10] S. J. Pan and Q. Yang. “A Survey on Transfer Learning”, *IEEE Transactions on Knowledge & Data Engineering*, vol.22, pp.1345-1359, 2010.

[11] C. Fernando, D. Banarse, C. Blundell, Y. Zwols, D. Ha, A. Rusu, A. Pritzel, and D. Wierstra. “PathNet: Evolution Channels Gradient Descent in Super Neural Networks”, *arXiv:1701.08734*, 2017.

[12] M. Long and J. Wang. “Learning Transferable Features with Deep Adaptation Networks” *Proceedings of the 32nd International Conference on Machine Learning, PMLR*, vol.37, pp.97-105, 2015.
[13] M. Long, H. Zhu, J. Wang, and M.I. Jordan, “Unsupervised Domain Adaptation with Residual Transfer Networks”, *Advances in Neural Information Processing Systems (NIPS)*, 2016.
[14] S. Lee, D. Kim, N. Kim, and S. Jeong. “Drop to Adapt: Learning Discriminative Features for Unsupervised Domain Adaptation”, *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2020.
[15] H. Maâmatou, T. Chateau, S. Gazzah, Y. Goyat, and N.E. Amara. “Transductive Transfer Learning to Specialize a Generic Classifier Towards a Specific Scene”, *VISIGRAPP*, 2016.
[16] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Gool. “Domain Adaptive Faster R-CNN for Object Detection in the Wild”, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.3339-3348, 2018.
[17] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang. “A survey of transfer learning”. *Journal of Big Data*, vol.3, pp. 9, 2016.
[18] H. Zhang, W. Leng, X. Han, and W. Sun. “Category-oriented Adversarial Data Augmentation via Inline Similarity for Satellite Images”, *5th Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, China, 2022.
[19] J.O. Neill, “Learning to Avoid Negative Transfer in Few Shot Transfer Learning”, 2019.
[20] J. Ding, N. Xue, Y. Long, G. Xia, and Q. Lu. “Learning RoI Transformer for Detecting Oriented Objects in Aerial Images”. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2844-2853, 2019.
[21] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun. “Light-head R-CNN: In Defense of Two-stage Object Detector”. *arXiv preprint arXiv:1711.07264*, 2017.
[22] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, and H. Hu. “Deformable Convolutional Networks”, *IEEE/CVF International Conference on Computer Vision*, 2017.
[23] J. Ding, N. Xue, G. Xia, X. Bai, W. Yang, and M. Yang. “Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.44, pp.7778-7796, 2021.
[24] K. He, X. Zhang, S. Ren, and J. Sun. “Deep Residual Learning for Image Recognition”. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770–778, 2016.
[25] X. Yang, Q. Liu, J. Yan, Z. Feng, and T. He. “R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object”, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol.35, 2021.