

RESEARCH

Open Access



Learning from label proportions for SAR image classification

Yongke Ding¹ , Yuanxiang Li^{2*} and Wenxian Yu¹

Abstract

Synthetic aperture radar (SAR) image classification plays a key role in SAR interpretation. Due to the cost and difficulty of truth labeling for SAR images, the newly labeled samples available for image classification are very limited. This paper focuses on defining a new sample labeling method to solve the problem of truth acquisition for training data in SAR image classification. An efficient classification framework for high-resolution SAR images is presented in this paper, which is built on learning from uncertain labels. We use grid labeling for rapid training data acquisition by assigning a label to a group of neighboring pixels at a time. A novel SVM-based learning model is proposed to optimize the uncertain training data within the constraints of label proportions in each group and then predict the label of each sample for the test data based on the optimized training set. This work intends to explore a rapid labeling method called grid labeling for efficient training set definition and apply it to large-scale SAR image classification. The model demonstrates good performance in both accuracy and efficiency for scene interpretation of high-resolution SAR images.

Keywords: SAR, Image classification, SVM, Label proportion, Land cover

1 Introduction

Synthetic aperture radar (SAR) image classification or land use/land cover (LULC) mapping plays an important role in many and diverse SAR applications [1]. Supervised learning is one of the most popular methods for SAR image classification. Supervised classification methods usually take a number of labeled samples to train the classifiers, and the classification accuracy depends on the selected training samples [2]. The acquisition of the ground truth for training data is expensive and time consuming. This is especially true for massive, high-resolution (HR) SAR data, such as TerraSAR-X, COSMO-SkyMed, etc. Due to the SAR imaging mechanism, the complex backscattering effects of objects and scenes in such HR SAR images will lead to local noises, double reflection, and even object deformation and part missing, making sample labeling highly difficult. So, researchers need to label the training samples more precisely and more carefully and even need the support of professional interpreters, field surveys, or other relevant information,

such as ancillary multispectral, geographic information system (GIS), or elevation data [3, 4]. The cost and difficulty of sample labeling have highly restricted the applications of large amounts of available HR SAR data, even though the data contains rich details of objects and scenes.

Related works on efficient training set definition for remote sensing image classification are mainly based on semisupervised learning with limited or small training sets, including transfer learning (TL) and active learning (AL). TL-based classification methods are designed to achieve high classification accuracy with relatively small number of labeled samples from the new image (target domain) by efficient reuse of the training data from the previous different but relative images (source domain) [5–7]. And AL-based classification methods focus on reducing the number of samples to be labeled by a human expert through iteratively choosing the most informative (i.e., uncertain and diverse) samples from the target domain [6–8]. A common practice is to combine TL and AL in the same learning framework [6, 7]. The abovementioned approaches mainly focus on effective and efficient sample selection from the same and/or different sources of images to reduce the labeling cost. The quality of pre-

*Correspondence: yuanxli@sjtu.edu.cn

¹Center for Advanced Sensing Technology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

Full list of author information is available at the end of the article

viously labeled sample will affect the classification result. And labeling for new samples is still difficult and inefficient as it is performed both in pixel-level labeling and in an iterative way. This paper tends to solve the problem of sample truth acquisition for SAR classification by exploring a new sample labeling method and the corresponding learning method.

The issue of learning from label proportions is raising attentions in the machine learning area [9–12]. For learning from label proportions, the training samples are divided into groups and label proportions of samples in each group are given as sample truth, instead of giving the label of each sample in the training set [11]. Learning from label proportion makes it possible to predict the label of each sample by labeling the training samples group by group, which shows potential for efficient classification of HR SAR images.

However, current classification methods based on label proportions [10, 11] usually come from the machine learning area, which are not suitable for SAR image classification because (1) these methods are based on detailed label proportions, that is, requiring proportions for each label in a group, which is still inefficient for sample labeling; and (2) these methods cannot perform multiclass classification directly. Although they can be extended to multiclass classification [11, 13], this also means extra computing.

The aim of this paper is to provide a unified framework with efficient sample labeling and model learning for SAR images. Inspired by the idea of learning from label proportions, we firstly introduce grid labeling to get the truth of training data more efficiently by giving the proportions of the labels in each group. Then, we present a novel support vector machine (SVM)-based learning model to eliminate the negative effects of label uncertainty from grid labeling with the support of label proportions. Moreover, an efficient inference method is proposed to approximately optimize the learning model.

The main contributions of our work are (1) introducing grid labeling for SAR image classification, to acquire the truth of the training data rapidly and at low cost; (2) building an SVM-based learning model by taking in consideration the constraints of label proportions; and (3) proposing an efficient inference method to reweight samples according to sample reliability and label proportions. Incorporation of label proportions and reweighting strategy helps to enhance the robustness of the SVM classifier with the existing uncertain (partially mislabeled) samples from grid labeling. To evaluate our method, we consider HR TerraSAR-X images of Tianjin and Rosenheim with four LULC classes [1]: urban area (UA), woodland (WL), open area (OA, such as farmland, grassland, bare soil, etc.), and water body (WB). We also have developed a software tool to implement grid labeling. First, our method outperforms

the state-of-the-art approach, i.e., α SVM in [11]. Besides, it demonstrates closely comparable accuracy with traditional supervised classification methods working with accurate pixel labeling but greatly reduces the labeling cost of training data, making large-scale SAR image classification more realizable.

This paper is organized as follows. Section 2 presents the SAR image classification framework based on grid labeling. Section 3 introduces grid labeling concept for efficient training set definition. Section 4 describes the proposed learning model with label proportions, including an efficient approximate inference algorithm based on alternating optimization strategy. Section 5 reports experimental results on the TerraSAR-X datasets and a simulated dataset. Our work is concluded in Section 6.

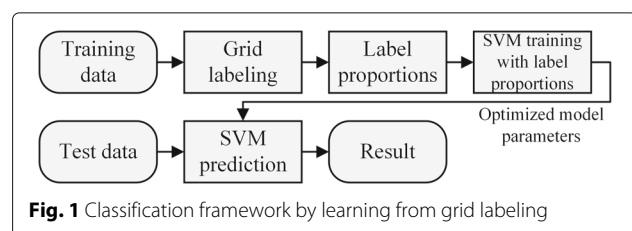
2 The classification framework

The proposed SAR image classification framework by learning from grid labeling is illustrated in Fig. 1. It combines boosting-based SVM and label proportions for efficient learning from uncertain labels. At the stage of grid labeling for training data, we slice the original SAR image into grid regions of 200×200 pixels, similar as in [14]. Such a grid region is called a “cell” in this paper. We select a subset (part) of all the cells as the training set T and assign a single label to each cell C in T , which is shared by all the pixels in the cell. So grid labeling outputs label proportions of each cell in the training data. Then at the stage of training, the SVM classifier [15, 16] with boosting strategy [17] is adopted to optimize the training data by reweighting the training samples within the constraints of label proportions. Finally, the SVM classifier predicts the label for each sample in the test data based on the optimized model parameters.

3 Grid labeling for rapid training set definition

3.1 Grid labeling concept

Grid labeling (GL) is introduced in this paper to acquire training data with high efficiency and at low cost. Instead of assigning one label to each pixel in the training set as in pixel labeling (PL), we assign a single label to each cell by grid labeling. All pixels in the cell will share the same label. Assume a proportion of the cells obtained by slicing the original SAR images are selected randomly and the training set is made up of pixels in these cells. The label for a cell C is determined by the major class with the maximum



number of pixels in C . We call it “naive grid labeling” as it does not provide any other information than the label of the major class in a cell.

Comparison of grid labeling and pixel labeling is illustrated in Fig. 2. For pixel labeling in Fig. 2c, the human expert carefully assigns a label for each pixel in the image, while for grid labeling, he or she just needs to estimate the major class in a cell and assigns the corresponding label to all the pixels in the cell, as in Fig. 2b. Obviously, grid labeling will introduce label uncertainty into the training set, for the pixels with other classes than the major one are still assigned the major label forcefully, e.g., on the boundary of WL and WB in Fig. 2b. But the potential of grid labeling in practical large-scale SAR image understanding makes it valuable to explore classification methods based on learning from grid labeling.

Although we usually accomplish pixel labeling by drawing a polygon on the image to annotate a region composed of pixels with the same class and fill it with a certain color, several factors will still make pixel labeling a hard and laborious task, such as various scales and shapes of the polygons, the unclear criterion how small a region can be ignored, and some regions with uncertain labels. Compared to traditional pixel labeling, the main advantage of grid labeling is its efficiency and operability, which includes the following: (1) Low cost and high efficiency. Grid labeling means greatly reduced cost and high efficiency, making large-scale SAR image classification more realizable; (2) Significant reduction of the difficulty in labeling HR SAR images. The backscattering effects of objects and scenes in HR SAR images are complex, making it difficult to precisely assign a label for each pixel. For grid labeling, the person in charge of labeling needs no special training and

avoids expensive field surveys in the labeling work; (3) Grid labeling also emphasizes more context clues than pixel labeling, which means it is more suitable for scene understanding.

3.2 Proportional grid labeling

As mentioned above, naive grid labeling performs on the groups (grid cells) from directly slicing the images in spatial domain, which is an efficient and convenient way for sample labeling. But naive grid labeling provide too little information (only the label of the major class in a cell) on sample classes, which will inevitably introduce label uncertainty and eventually reduce the classification accuracy. According to the idea of learning with label proportions [9–11], we can learn a model to predict labels of the individual samples by grouping the training samples and providing proportions of the labels in each group. However, the current definition of label proportion [11] ignores spatial sources of samples in each group, which means samples in a group are not necessarily from the same local region, making it not convenient for sample labeling of remote sensing images. Thus, we introduce proportional grid labeling (proportional GL) by combining grid labeling with label proportions to implement rapid training set definition for SAR images. By estimating the proportions of each label in cell C , proportional GL not only can implement convenient and efficient sample labeling for SAR images as naive grid labeling but also can provide more information on sample classes.

Let $\mathcal{L} = \{l_i\}_{i=1}^M$ be the label set, where M is the number of labels in \mathcal{L} . Assume the training set $T = \{x_i\}_{i=1}^N$ contains N training samples x_i . Then, T can be divided into K disjoint cells $C_k (k=1, \dots, K)$ where $\bigcup_{k=1}^K C_k = \{1, \dots, N\}$, and $\{x_i | i \in C_k\}$ is the set of samples in C_k . Label proportion is defined as the proportion of samples with a certain

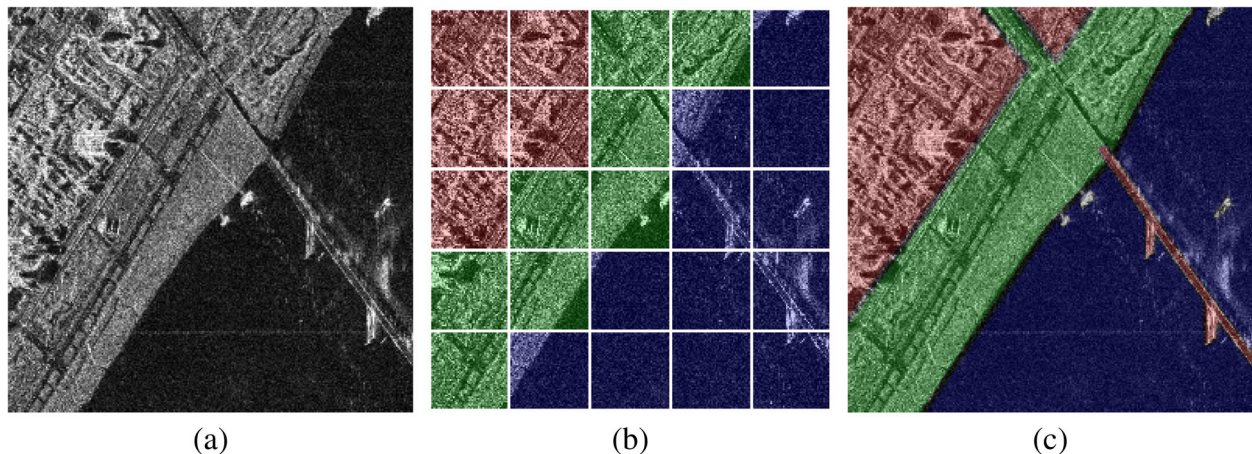


Fig. 2 Grid labeling vs. pixel labeling. **a** Original SAR image region. **b** Labeling the SAR image region in grid of 5×5 cells. **c** Labeling the SAR image region in pixel. In **b** and **c**, we use red for UA, green for WL, and blue for WB, as an overlay on the SAR image region

label in a grid cell. The label proportion for $l_i \in \mathcal{L}$ in cell C_k is

$$p_k(l_i) = \frac{1}{|C_k|} \sum_{j \in C_k} [y_j^* = l_i], \quad (1)$$

where y_j^* is the real pixel-level ground truth of x_j , $|C_k|$ denotes the total number of samples in cell C_k . Then, proportional GL can be presented in three forms by giving label proportions for each class or the major class in a cell during the sample labeling process,

1) Multiclass GL: The task of multiclass GL is to estimate the proportion $p_k(l_i)$ for each $l_i \in \mathcal{L} (i=1, \dots, M)$ in a grid cell C_k by a human expert.

2) Majorclass GL: Based on the definition of multiclass GL, assume $l^{(k)} = \arg \max_{l_i \in \mathcal{L}} p_k(l_i)$ is the label which has the largest (or major) proportion in C_k and $p_k(l^{(k)})$ is its proportion. Then, $l^{(k)}$ is called the major class of cell C_k . Using p_k to denote $p_k(l^{(k)})$ for short, we get,

$$p_k(l^{(k)}) = p_k. \quad (2)$$

$$p_k(\mathcal{L} \setminus l^{(k)}) = 1 - p_k, \quad (3)$$

where $\mathcal{L} \setminus l^{(k)} = \{l_j \in \mathcal{L} | l_j \neq l^{(k)}\}$ is the set of all labels except $l^{(k)}$. So the task of majorclass GL is just to give p_k for major class $l^{(k)}$ in a grid cell C_k by a human expert.

3) Naive GL: Based on the definition of majorclass GL, we can predefine $p_k = 1$ for simplicity, which also means $p_k(\mathcal{L} \setminus l^{(k)}) = 0$. Then, the human expert only needs to indicate the majorclass label $l^{(k)}$ in a grid cell C_k , without estimating the proportion of any label. In this way, majorclass GL is indeed degraded to the naive grid labeling (naive GL) described in Section 3.1.

4 Learning from grid labeling

In this section, we focus on exploring SAR image classification method by learning from majorclass GL. The main reason we choose majorclass GL for sample labeling relies on its good balance between efficiency on sample labeling and integration with label priors. Majorclass GL is integrated with label proportions of the major class in a cell, which is the most important one among all labels. Although multiclass GL provides more information than majorclass GL, it is much more difficult to estimate the proportion for each class than the major class in sample labeling. Besides, the optimization algorithm has to satisfy the proportion of each label in the label set for each cell in order to solve a learning problem under the constraint of multiclass proportions. However, it only needs

to satisfy the proportion of the majorclass label for each cell to solve a learning problem based on majorclass proportions. So it is much more difficult to solve a learning problem under the constraint of multiclass proportions than the majorclass proportions.

4.1 The learning model with label proportions

We build a learning model with label proportions based on the SVM classifier [16] with boosting strategy [17]. The SVM classifier takes the simple but efficient features (i.e., backscattering intensity, texture, and supertexture) proposed in [1] as its inputs and outputs the land cover types of the SAR images. Optimization of the training set from grid labeling is implemented by reweighting strategy of the boosting algorithm [17, 18], which has been improved by us through introducing a new penalty term for label proportions.

We formulate the SVM-based model under the large-margin framework as below [7, 11],

$$\begin{aligned} \min_{\mathbf{w}, b, \omega} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + \lambda \sum_{i=1}^N \omega_i L(y_i, \mathbf{w}^T \varphi(x_i) + b) + \lambda_p \sum_{k=1}^K L_p(\tilde{p}_k(\omega), p_k) \\ \text{s.t. } & \omega_i \in [0, 1], i = 1, \dots, N \end{aligned} \quad (4)$$

where ω_i is the weight for training sample x_i ; $\omega = [\omega_1, \dots, \omega_N]^T$ is the weight vector; y_i is the true label of x_i from grid labeling; $L(\cdot) \geq 0$ is the loss function for the traditional SVM model; another loss function $L_p(\cdot) \geq 0$ is designed to penalize the dissimilarity between the true label proportion p_k and the predicted label proportion \tilde{p}_k based on ω ; we define $\tilde{p}_k(\omega) = \frac{1}{|C_k|} \sum_{i \in C_k} [\omega_i > 0]$ because

a nonzero value will be assigned to ω_i if the majorclass label $l^{(k)}$ is reliable to be the true label of x_i ($i \in C_k$), i.e., $y_i = l^{(k)}$; $\varphi(x_i)$ maps x_i into a higher-dimensional space; λ and λ_p are the regularization parameters. Our task is to simultaneously optimize the weight ω and the SVM model parameters \mathbf{w} and b to ensure \tilde{p}_k estimated using ω is close enough to p_k .

Based on pixel-level training and label proportion constraints from grid labeling, this model not only can implement pixel-level classification for SAR images but also can improve the efficiency of sample labeling and reduce the corresponding cost. We build this model based on SVM because the training set can be optimized by reweighting the importance of each sample used for training in SVM model. This strategy has been adopted in transfer learning methods [7, 17]. And we can easily extend the number of classes for classification by defining a training set with rich classes for SVM classifier.

4.2 Model inference using LpcSVM algorithm

The unknown sample weight ω can be seen as a link between supervised learning loss $L(\cdot)$ and label proportion loss $L_p(\cdot)$ in Eq. (4). So optimization for Eq. (4) can be solved by alternating optimization strategy [11, 19], i.e., optimizing (\mathbf{w}, b) and ω one at a time by fixing the others. For fixed sample weights $\omega = [\omega_1, \dots, \omega_N]^T$, the optimization of Eq. (4) w.r.t \mathbf{w} and b becomes a classic SVM problem. When \mathbf{w} and b are fixed, ω can be updated according to label agreement and label proportion constraints. The baseline strategy of weight distribution update is to assign lower weights for misclassified or unreliable samples in the training set [20]. Assume the majority of samples in the training set are correctly labeled by grid labeling. So correctly classified (reliable) training samples are associated to strong weights and misclassified samples are assigned with weak weights in the reweighting process. An updated ω means an updated classifier, which is expected to be more trustworthy than previous ones, since it is trained with more reliable samples.

The alternating optimization procedure of Eq. (4) is summarized in Algorithm 1, which is called label-proportion-constrained SVM (LpcSVM) algorithm. By initializing $\omega_i = 1 (i = 1, \dots, N)$, we obtain \mathbf{w} and b through the traditional SVM training,

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \lambda \sum_{i=1}^N \omega_i L(y_i, \mathbf{w}^T \varphi(x_i) + b) \quad (5)$$

Algorithm 1 LpcSVM: label-proportion-constrained SVM

Initialize the sample weight $\omega_i = 1 (i=1, \dots, N)$, sample label $y_i = l^{(k)} (i \in C_k)$ and the maximum number of iterations nT .

for $t = 1, \dots, nT$ **do**

1) Fix ω_i to solve \mathbf{w} and b using the traditional SVM training in Eq. (5).

2) Fix \mathbf{w} and b to solve ω_i using cell-wise optimization in Eq. (7).

Prepare the posterior distribution $P(l|x_i)$ for the training data using LIBSVM based on \mathbf{w} and b .

for $k = 1, \dots, K$ **do**

1) Calculate $R(l^{(k)}|x_i)$ to measure sample reliability using Eq. (8).

2) Sort $x_i (i \in C_k)$ in ascending order by $R(l^{(k)}|x_i)$.

3) Update sample weight $\omega_i (i \in C_k)$ using Eq. (9).

end for

end for

SVM prediction for the test data using the optimized model parameters.

In Eq. (5), we set $y_i = l^{(k)}$ for $x_i (i \in C_k)$ as the major class is used for labeling all the samples in C_k during the training process. When \mathbf{w} and b are obtained, the problem of Eq. (4) becomes

$$\begin{aligned} \min_{\omega} \sum_{i=1}^N \omega_i L(y_i, \mathbf{w}^T \varphi(x_i) + b) + \frac{\lambda_p}{\lambda} \sum_{k=1}^K L_p(\tilde{p}_k(\omega), p_k) \\ \text{s.t. } \omega_i \in [0, 1], i = 1, \dots, N \end{aligned} \quad (6)$$

The problem of Eq. (6) can be solved by optimizing $\omega_i (i \in C_k)$ on each cell separately as the influence of each cell C_k on the objective function in Eq. (6) is independent. This means Eq. (6) can be converted into a cell-wise optimization form,

$$\begin{aligned} \min_{\{\omega_i | i \in C_k\}} \sum_{i \in C_k} \omega_i L(y_i, \mathbf{w}^T \varphi(x_i) + b) + \frac{\lambda_p}{\lambda} L_p(\tilde{p}_k(\omega), p_k) \\ \text{s.t. } \omega_i \in [0, 1], i \in C_k \end{aligned} \quad (7)$$

where the absolute loss $L_p(\tilde{p}_k(\omega), p_k) = |\tilde{p}_k(\omega) - p_k|$ is used in this paper. We adopt an approximate cell-wise optimization method for estimating weight ω_i in Eq. (7).

First, we calculate potential $R(l^{(k)}|x_i)$, a measure for sample reliability or label agreement, that is, how likely x_i will be classified as $y_i = l^{(k)}$ by the classifier. It is defined as

$$R(l^{(k)}|x_i) = E(l^{(k)}|x_i) - \min_{l \in \mathcal{L} \setminus l^{(k)}} (E(l|x_i)), \quad (8)$$

where $E(l|x_i) = -\log(P(l|x_i))$. The class posterior distribution $P(l|x_i)$ is estimated by LIBSVM [16]. LIBSVM uses the method in [21] to estimate class probability for classification problem. The basic idea of the method is to estimate pairwise (i.e., one-against-one) class probability from decision values (or cost values) and then solve an optimization problem to get multi-class probability. Compared to the simple potential form $E(l^{(k)}|x_i)$, the definition in Eq. (8) is expected to better depict the potential of x_i to be classified as $l^{(k)}$, by measuring the distance between the potential of label $l^{(k)}$ and the minimal value among all the other labels (i.e., the label most likely to replace $l^{(k)}$). Smaller value of $R(l^{(k)}|x_i)$ means $l^{(k)}$ is more reliable to reflect the true label of x_i .

Then, we sort $x_i (i \in C_k)$ in ascending order by $R(l^{(k)}|x_i)$ and note the sorted sample sequence as x_{δ_i} , where δ_i is the index value (or position) of x_i in the sorted sequence.

Next, update sample weight for $x_i (i \in C_k)$ using the following reweighting formula,

$$\omega_i = \begin{cases} 1, \delta_i \leq N_m \\ \exp\left(-\frac{(\delta_i - N_m)^2}{\theta |C_k|^2}\right), N_m < \delta_i \leq N_s \\ 0, \delta_i > N_s \end{cases} \quad (9)$$

where $N_s = \text{int}(p_k |C_k|)$ is the sample number of major class l_k in C_k ; $N_m = \frac{|C_k|}{M}$ is the number of samples to

be assigned with weight 1; and θ is a parameter to control the degree of weight weakening for $N_m < \delta_i \leq N_s$. Typically, $\theta = 0.5$ will work well in most cases. According to the formula, we assign strong weight ($\omega_i > 0$) for the first N_s samples with smallest $R(l^{(k)}|x_i)$ (where i satisfies $\delta_i \leq N_s$) and zero weight for the others. To further enhance the robustness of our method to the estimation error of label proportion p_k , the sample weight is designed to introduce a penalty term for $N_m < \delta_i \leq N_s$, instead of simply assigning $\omega_i = 1$ for all $\delta_i \leq N_s$, as illustrated in Fig. 3.

In Eq. (9), N_s is used to control the difference between the predicted and true label proportions by setting the number of the predicted majorclass samples to that of the true majorclass samples in C_k explicitly. According to $\tilde{p}_k(\omega)$ defined before, we have

$$\tilde{p}_k(\omega) = \frac{1}{|C_k|} \sum_{i \in C_k} [\omega_i > 0] = \frac{N_s}{|C_k|}. \quad (10)$$

Define the operator “int” for rounding a float to a smaller integer. By setting $N_s = \text{int}(p_k |C_k|)$, we get $\tilde{p}_k(\omega) \approx p_k$. Then $L_p = |\tilde{p}_k(\omega) - p_k| \approx 0$, which means the loss function in Eq. (7) is minimized to its first term, i.e., the weighted sum of the loss values of the chosen N_s samples. So we have

$$\begin{aligned} \min_{\{\omega_i | i \in C_k\}} \sum_{i \in C_k} \omega_i L(x_i) + \frac{\lambda_p}{\lambda} L_p(\tilde{p}_k(\omega), p_k) \\ = \min_{\{\omega_i | i \in C_k\}} \sum_{i \in C_k} \omega_i L(x_i) \end{aligned} \quad (11)$$

where we use $L(x_i)$ to denote $L(y_i, \mathbf{w}^T \varphi(x_i) + b)$ for short.

Let ω'_i be the updated version of sample weights ω_i using the reweighting strategy in Eq. (9). Then, ω'_i is decreasing (non-increasing) with δ_i (see Fig. 3). Recall that $\{x_{\delta_i}\}$ is the sample sequence after sorting $\{x_i\}$ in ascending order of the potential $R(l^{(k)}, x_{\delta_i})$, which means $R(l^{(k)}, x_{\delta_i})$ is increasing (non-decreasing) with δ_i . For a sample x_{δ_i} , a smaller value of the potential $R(l^{(k)}, x_{\delta_i})$ indicates a greater probability of $l^{(k)}$ to be the true label of x_{δ_i} , which

also means a smaller value of sample cost $L(x_{\delta_i})$ in Eq. (7). So $L(x_{\delta_i})$ is also increasing (non-decreasing) with δ_i .

According to the principle of rearrangement inequality [22], $\sum_{i \in C_k} \omega_i L(x_i)$ will be minimized when the sequences $\{\omega_i\}$ and $\{L(x_i)\}$ are in opposite orders. Since $\{\omega'_i\}$ is in descending order by δ_i and $\{L(x_{\delta_i})\}$ is in ascending order by δ_i , we have $\min_{\{\omega_i | i \in C_k\}} \sum_{i \in C_k} \omega_i L(x_i) = \sum_{i \in C_k} \omega'_i L(x_{\delta_i})$. Finally, Eq. (7) is minimized as

$$\begin{aligned} \min_{\{\omega_i | i \in C_k\}} \sum_{i \in C_k} \omega_i L(x_i) + \frac{\lambda_p}{\lambda} L_p(\tilde{p}_k(\omega), p_k) \\ = \sum_{i \in C_k} \omega'_i L(x_{\delta_i}) \end{aligned} \quad (12)$$

which means the cost function in Eq. (7) is minimized by setting higher weight ω'_i to lower sample cost $L(x_{\delta_i})$ using our LpcSVM algorithm.

Besides, we choose $N_m = \frac{|C_k|}{M}$ in Eq. (7) for the following reason. As there are M labels for classification, the infimum of p_k is $\inf(p_k) = \frac{1}{M}$. We get $\inf(N_s) = \inf(p_k) \times |C_k| = \frac{|C_k|}{M}$. Despite possible noises in p_k , at least $N_m = \frac{|C_k|}{M}$ samples in each cell should be made full use of by assigning weight 1 to them. A larger value of δ_i means a less reliable sample. In our reweighting strategy, weakened weights are assigned to less reliable ones (i.e., $N_m < \delta_i \leq N_s$) among the N_s samples chosen for the next iteration (see Fig. 3). Then, small fluctuation of N_s will not cause considerable changes in the weight distribution, making the method of better robustness to possible estimation errors in label proportions. Experimental performance of the robustness of our method to the estimation error in label proportions will be described in Section 5.4.

By combining label agreement [17] and label proportion similarity between the predicted and true labels, our reweighting strategy will not only weaken the role of unreliable samples in the training set T but also keep the important role of a certain number of reliable samples within the constraints of label proportions for the major class in each cell.

The updated ω is used to retrain the SVM classifier. After a number of iterations, strong weights will be assigned to the reliable samples and weak weights to the others, producing an optimized pixel-level training set (in form of optimized sample weights). Finally, the classifier will predict the label for each sample in the test data using the optimized model parameters.

Although partially inspired by the idea of α SVM [11], our model is essentially different with α SVM in the following respects: (1) α SVM performs on multiclass label proportions, while our model performs on majorclass label proportion, which makes sample labeling more efficient; (2) α SVM only deals with binary classification,

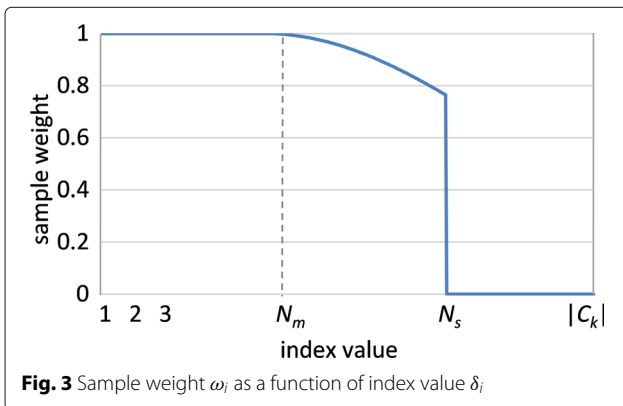


Fig. 3 Sample weight ω_i as a function of index value δ_i

while our model deals directly with multi-class classification; (3) Besides, our model is optimized based on sample reweighting, which is totally different with α SVM and also makes multiclass classification realizable.

5 Experiments

5.1 Dataset description and experiment setting

We use two full-scene HR TerraSAR-X images of Tianjin and Rosenheim areas (described in Table 1) to evaluate the performance of our method. For classification, we consider four classes for Tianjin and Rosenheim areas as in [1], i.e., UA, WL, OA, and WB. We use about 10% of all the cells coming from slicing the original SAR images for training the SVM model by random selection strategy and the whole images for classification using SVM, which is a common practice in remote sensing image classification [1, 23]. The number of training samples for each label is proportional to the label distributions in the image to avoid class imbalance. All the reported results represent average over ten random realizations of the training set for each image.

The SVM classifier is configured with radial basis function (RBF) kernel and probability estimation [16]. In our experiment, the parameter λ in Eq. (7) is tuned from $\{0.1, 1.0, 10\}$ and we found $\lambda=1.0$ works better for our test. The parameter λ_p in Eq. (7) is tuned from $\{\lambda, 10\lambda, 100\lambda, 1000\lambda\}$. A larger value of λ_p means introducing more penalty from label proportions. Theoretically, the parameter θ in Eq. (9) controls the penalty on the less reliable training samples with $N_m < \delta_i \leq N_s$. A larger θ means assigning larger weights to the corresponding samples. In our test, we tune θ from $\{0.3, 0.4, 0.5, 0.6, 0.7\}$ and found $\theta=0.5$ works better for our experiment. We use efficient low-level features such as backscattering intensity, texture, and supertexture [1] for the SVM classifier to take as input. The features are defined as follows.

The intensity feature f_i^I is the backscattering intensity value at pixel i . Suppose we define a patch i centered at pixel i , which is actually a local region at pixel i for texture measuring. Both texture and supertexture features are extracted on patches. The texture feature f_i^{tex} represents the coefficient of variation in a patch, that is, $f_i^{tex} = \sigma_i^{tex} / \mu_i^{tex}$, where μ_i^{tex} gives the mean backscattering intensity value and σ_i^{tex} represents the standard deviation of the backscattering intensity for all the pixels in patch i . In our test, we adopt a patch of 11×11 pixels, which

will cover a block with homogeneous texture (e.g., a residential building) under the resolution of our test images (around 1.5 m). The supertexture feature f_i^{sup} measures the similarity of textures between neighboring patches, thus providing clues for texture context of neighboring patches. For patch i , supertexture is defined as

$$f_i^{sup} = \frac{\sigma_i^{sup}}{\mu_i^{sup}} = \sqrt{\frac{1}{n} \sum_{j \in N(i)} (f_j^{tex} - \mu_i^{sup})^2} / \frac{1}{n} \sum_{j \in N(i)} f_j^{tex} \quad (13)$$

where μ_i^{sup} is the average of f_j^{tex} and σ_i^{sup} is the standard deviation of f_j^{tex} in the neighborhood $N(i)$ of patch i , n is the number of patches in $N(i)$. A 5×5 rectangular neighborhood is exploited in our work for an appropriate measurement of scene homogeneity and texture context (such as a piece of residential area or green field). Then, the feature vector $f_i = [f_i^I, f_i^{tex}, f_i^{sup}]^T$ for each pixel x_i is taken as input for the SVM classifier.

We have designed four groups of experiments to test our method from different aspects. The first is to show overall experiment results by comparing our method to the baseline methods. The second is to analyze the classification performance of our method with cell size and iteration varying. The third is to explore the robustness of our method to p_k estimated in majorclass GL. The last is to further apply our method to classification of a simulated dataset. Besides, we also discuss the efficiency and cost advantages of our method over traditional methods at the end of this section.

5.2 Overall experimental results

We compared our method (label-proportion-constrained SVM with grid labeling, GL+LpcSVM) with the traditional SVM classifier with pixel labeling (PL+SVM), the traditional SVM classifier with naive grid labeling (GL+SVM), and α SVM, which is also based on learning from label proportions. For the implementation of α SVM, we choose alter- α SVM as our baseline method instead of conv- α SVM for the former performs better than the latter in most situations [11]. The original α SVM is for binary classification. We adopt the method similar to [13] to extend α SVM for multiclass classification. The main idea of the method is to use pairwise label switching between classes to improve the objective function for multi-class classification. The classification accuracies along with Kappa coefficient for Tianjin and Rosenheim based on these methods are listed in Table 2. The overall accuracy is calculated as the ratio of the number of correctly classified samples to the total number of samples in the image.

As can be seen from Table 2, GL+SVM does not perform so well due to the uncertain labels introduced by

Table 1 Datasets used to test the methods

Site	Imaging mode	Pixel posting	Size
Tianjin	Stripmap	2.0 m	8911×8787
Rosenheim	Spotlight	1.25 m	9504×8330

Table 2 Overall classification accuracy for the SAR images (%)

Method	Tianjin		Rosenheim	
	Overall accuracy	Kappa coefficient	Overall accuracy	Kappa coefficient
PL+SVM	87.41 \pm 0.43	0.7762 \pm 0.0132	68.17 \pm 0.59	0.5880 \pm 0.0091
GL+SVM	83.13 \pm 0.47	0.6698 \pm 0.0104	64.52 \pm 0.61	0.4623 \pm 0.0324
α SVM	84.86 \pm 0.53	0.7144 \pm 0.0121	65.92 \pm 0.95	0.4847 \pm 0.0083
GL+LpcSVM with $n_k = 0$	86.38 \pm 0.68	0.7278 \pm 0.0131	67.34 \pm 0.55	0.5728 \pm 0.0079
GL+LpcSVM with $n_k \sim N(0, 0.05^2)$	86.33 \pm 0.74	0.7226 \pm 0.0143	67.14 \pm 0.53	0.5719 \pm 0.0077

grid labeling. But α SVM performs better than GL+SVM through learning from label proportions. And our method (GL+LpcSVM with and without noise n_k in p_k) outperforms α SVM in classification accuracy with about 1.5%. Moreover, α SVM cannot deal with multiclass classification directly, which means extra computing will be needed for α SVM to deal with multiclass classification. And α SVM is based on detailed label proportions, which also means more cost than LpcSVM (based on majorclass GL) in sample labeling. Although it is unfair to compare our method with PL+SVM for the latter is based on the accuracy training set from pixel labeling, we still take PL+SVM as a baseline method. GL+LpcSVM (with and without noise n_k in p_k) demonstrates comparable classification accuracy and Kappa coefficient with PL+SVM, while reducing labeling cost significantly by adopting grid labeling.

The classification results of Tianjin and Rosenheim areas using our method and the baseline methods are presented in Fig. 4. The ground truth maps are labeled manually by referring to the corresponding optical images, with undefined areas in black color. For all methods, no postprocessing steps are added after classification using SVM. In the experiment, we adopt a cell size of 200×200 and perform four iterations for LpcSVM. The criteria for selecting cell size and iteration number will be explained in Section 5.3.

5.3 Classification performance with cell size and iteration varying

The proportion of all mislabeled samples (i.e., $p_{\text{mis}} = \frac{1}{K} \sum_{k=1}^K (1 - p_k)$) in the training set as a function of the cell size of grid labeling is illustrated in Fig. 5. As can be seen from the figure, a larger cell size in grid labeling basically means a larger proportion of incorrectly labeled (mislabeled) samples in the training set. Thus, the cell size can be used as a direct measure of the total proportion of mislabeled samples, which will influence the classification accuracy of the proposed LpcSVM classifier. Larger cell size means lower cost for sample labeling as the number of cells to be labeled is smaller, but it also means lower

classification accuracy as the training set contains more unreliable samples (i.e., higher mislabeled proportion). So a balance between labeling cost and classification accuracy should be found in the application of grid labeling. The balance can be reached by choosing the right cell size.

Classification accuracy using GL+LpcSVM for the two SAR images as a function of the iteration number of LpcSVM under different cell sizes is shown in Fig. 6. We can conclude that, for a certain cell size, the classification accuracy on the two SAR images increases with the iteration numbers of LpcSVM, remaining stable after four iterations. More than four iterations will not improve the classification accuracy obviously, even though not reducing the accuracy in our test. So about four iterations of LpcSVM is applied in our experiments. When the cell size changes from 1×1 to 200×200 , the curve of classification accuracy falls gradually, but it drops dramatically from the cell size of 300×300 . Consequently, we adopt a cell of about 200×200 pixels in the application of grid labeling to reach a balance between labeling cost and classification accuracy (as the results in Table 2 and Fig. 4).

5.4 Robustness to p_k estimated in majorclass GL

Since label proportions estimated by human in grid labeling may differ from the true label proportions, there possibly exist errors in label proportion p_k . As we have introduced a penalty of weights at the end of the sequence of N_s samples chosen for the next training in the reweighting formula of Eq. (9), small errors in p_k will not degrade the performance of our method. Influence of label proportion errors on classification accuracy for the two SAR images is presented in Fig. 7. The test is performed with a cell size of 200×200 and four iterations. We simulate the errors in label proportion by adding noise n_k of a normal distribution with mean $\mu=0$ and standard deviation $\sigma = 0.05 \sim 0.20$ to the existing label proportion p_k . Then, the noisy label proportion is $p_k = p_k + n_k$, where $n_k \sim N(\mu, \sigma^2)$. According to the three-sigma rule, the value of n_k will basically fall in $[\mu - 3\sigma, \mu + 3\sigma]$, e.g., $[-0.15, 0.15]$ for $\sigma=0.05$. As can be seen from the figure, our method still maintains a relative high level of classification accuracy under the influence of errors in label proportions.

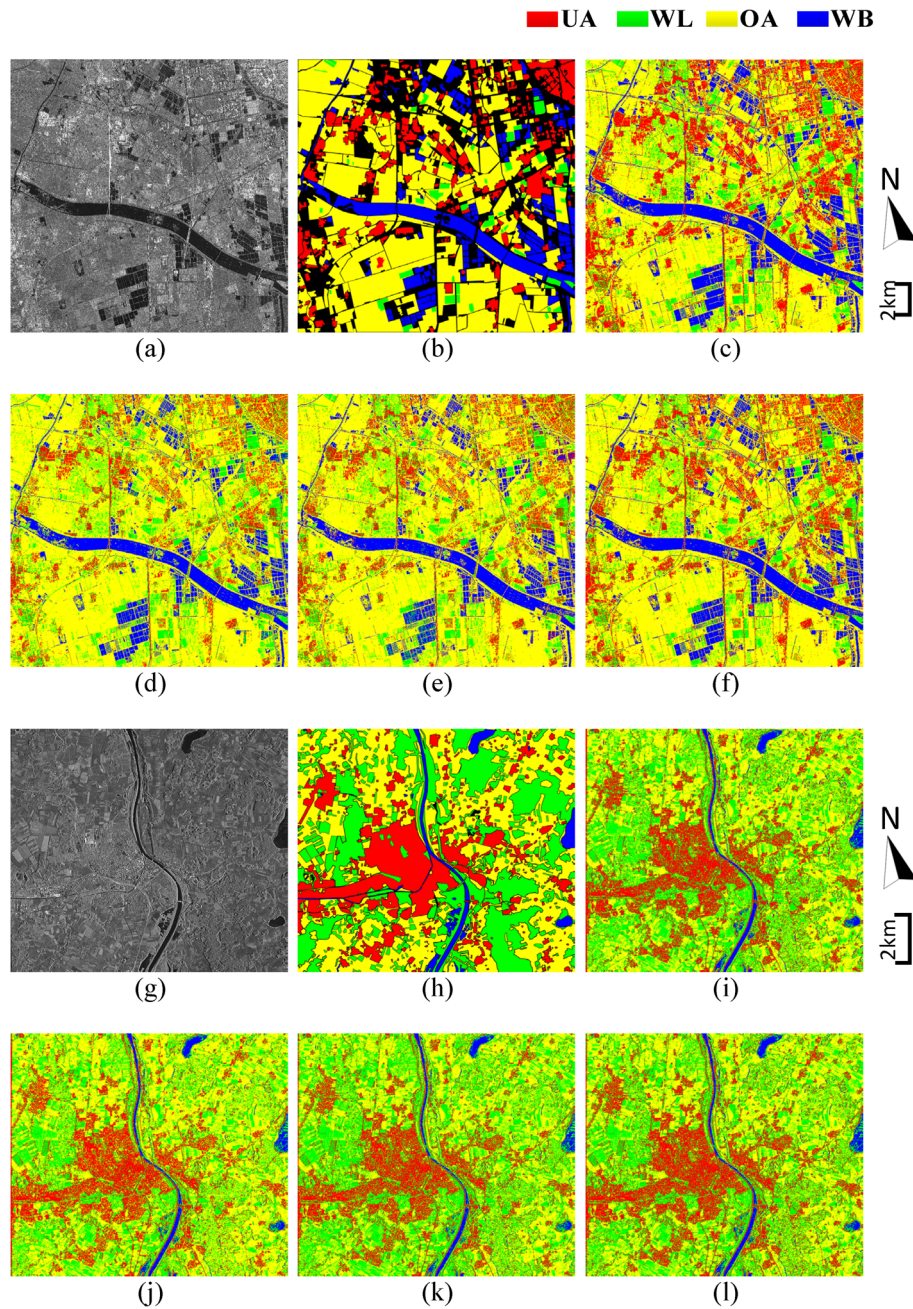


Fig. 4 Classification results for Tianjin and Rosenheim. **a** TerraSAR-X image of Tianjin. **g** TerraSAR-X image of Rosenheim. **b, h** Ground truth maps from pixel labeling (for accuracy calculation). **c, i** PL+SVM. **d, j** GL+SVM. **e, k** α SVM. **f, l** GL+LpcSVM with $n_k = 0$. Note that the result image of GL+LpcSVM with $n_k \sim N(0, 0.05^2)$ is not listed here for it looks very similar to that of GL+LpcSVM with $n_k = 0$

In order to estimate the standard deviation value of actual label proportion errors, grid labeling is performed on nearly 4000 cell regions from the two full-scene TerraSAR-X images used in the experiment. Five persons participate in the grid labeling stage, and each person labels each image twice. The order of cell regions in each image is randomly generated for each labeling by each person. So we obtain ten groups of different labelings

for the cell regions in each image. Based on our grid labeling activities, we found the standard deviation value of label proportion errors in grid labeling is basically around $\sigma=0.05$, which means the value of label proportion errors falls in the confidence interval $[-0.15, 0.15]$ with the probability of 99.7%, based on the three-sigma rule mentioned above. However, our method can still provide desirable classification results (85.92% for Tianjin and

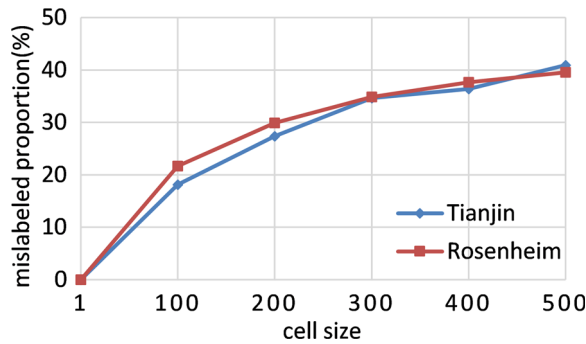


Fig. 5 Mislabeled proportion of training data varies with cell size. The mislabeled proportion of training data for Tianjin and Rosenheim is approximately proportional to the cell size of grid labeling. For cell size of 1×1 , grid labeling is actually degraded to pixel labeling

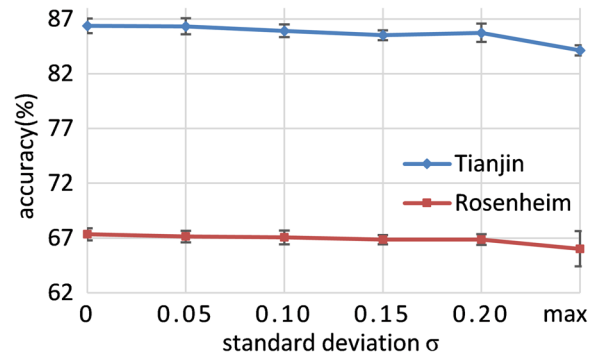
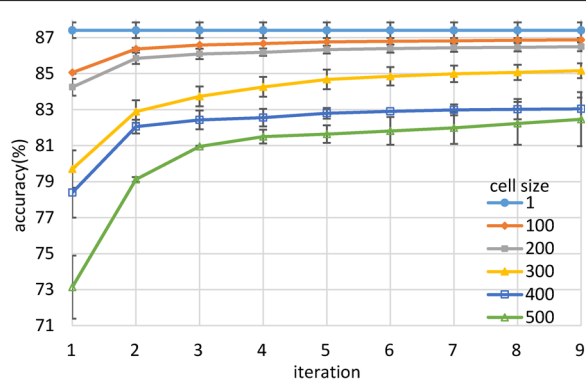
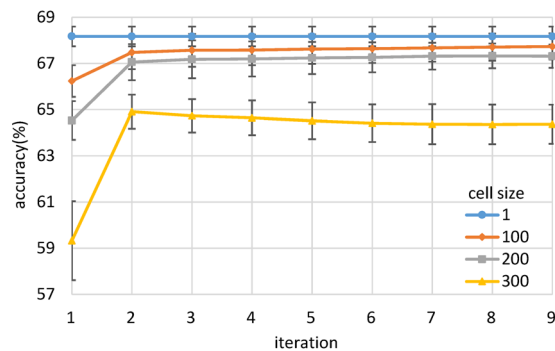


Fig. 7 Influence of label proportion errors on classification accuracy using GL+LpcSVM for SAR images of Tianjin and Rosenheim areas. Noise standard deviation σ ranges 0.05, 0.10, 0.15, and 0.20. Note that $\sigma = 0$ indicates $n_k = 0$, meaning no noises added in p_k . The special case that no label proportions are provided in the training data (degraded to naive GL) is also considered, denoted by $\sigma = \max$



(a)



(b)

Fig. 6 Classification accuracy using GL+LpcSVM for the SAR images of Tianjin (a) and Rosenheim (b) as a function of the iteration number of LpcSVM under different cell sizes. Cell size ranges $1 \times 1, 100 \times 100, \dots, 500 \times 500$. Note that cell size of 1×1 means the traditional SVM with pixel labeling (PL+SVM), making a horizontal reference line through different iterations. For Rosenheim, the maximum cell size is 300×300 as larger cell sizes will cause the land cover of water body (WB) disappear in the training set. Iterations of four will produce a stable value of classification accuracy

67.06% for Rosenheim) even under a noise with $\sigma = 0.10$ (see Fig. 7).

Besides, if no label proportions are provided (i.e., $p_k = 1, k=1, \dots, K$ as defined in Naive GL), our method can still work relatively well, outperforming GL+SVM about 1% in classification accuracy (see the case of $\sigma = \max$ in Fig. 7). In this condition, all the samples in cell C_k are selected for the next iteration (i.e., $N_s = |C_k|$), but they are reweighted differently according to the reliability of their labels, without the support of label proportions.

5.5 Classification with the simulated dataset

The dataset of a simulated SAR image is also used to further test our method. In order to simulate a SAR image of a large scene for classification, we adopt a widely used speckle-statistics-based simulation method as in Section 4.1.2 of [24]. Four types of homogeneous areas are considered to represent four land cover types in the simulated SAR image, and the pixels in each type of homogeneous areas follow a Rayleigh distribution. We implement each type of homogeneous area by setting the real and imaginary components of each pixel in the area to follow an independent and identical Gaussian (normal) distribution with zero mean and a variance σ^2 . Then, the amplitude A of the pixels follows a Rayleigh probability density $p(A) = \frac{A}{\sigma^2} \exp\left(-\frac{A^2}{2\sigma^2}\right), A \geq 0$ with its mean $\sigma\sqrt{\frac{\pi}{2}}$ and the variance $\frac{4-\pi}{2}\sigma^2$.

We set $\sigma=50, 110, 130, 150$ respectively to generate the four types of homogeneous areas (from dark to bright). The truth map of the TerraSAR-X image of Rosenheim is used for spatial configuration (such as locations, sizes, and shapes) of different homogeneous areas in the simulated image. Finally, we obtain a simulated image with size of

9504×8330, which contains four types of homogeneous areas to represent four different land cover types.

The experiment setting for the simulated image is the same with the two TerraSAR-X images, as mentioned in Section 5.1. The classification accuracies along with Kappa coefficient for the simulated image based on our method, and the baseline methods are shown in Table 3. We can see that our method (GL+LpcSVM) outperforms the baseline methods (GL+SVM and α SVM) in classification accuracy and Kappa coefficient. Our method also keeps a comparable performance with the traditional PL+SVM method.

5.6 Efficiency and cost advantages

As mentioned above, the classification accuracy of our method (GL+LpcSVM) is slightly lower than PL+SVM as the latter is based on the accuracy training set from pixel labeling. But we emphasize the efficiency and cost advantages of our method over PL+SVM. For efficiency measuring, we consider the total classification time t as the sum of training set definition time t_1 and time consuming t_2 of the SVM classifier. Even using polygon annotation, the efficiency of pixel labeling will still be very low due to various scales and shapes of the polygons, the unclear criterion how small a region can be ignored, and some regions with uncertain labels. It will be shown that the efficiency of classification can be improved significantly through the great reduction of training set definition time t_1 based on our method.

We have developed a software tool for proportional grid labeling to define the training set of classification, as shown in Fig. 8. The average time consuming for classification on the TerraSAR-X images of Tianjin and Rosenheim areas using the baseline PL+SVM, and our GL+LpcSVM is provided in Table 4. The test was conducted on an Intel i5-2520M dual core processor clocked at 2.50 GHz. For PL+SVM, t_1 is the time of pixel labeling, and for GL+LpcSVM, it is the time of grid labeling using the developed labeling tool. From the table, we can see that t_1 is much larger than t_2 for both methods. Our method makes the total classification time t reduced greatly by decreasing the training set definition time t_1 greatly and keeping the LpcSVM inference algorithm reaching stable within just three or four iterations (see Fig. 6), which means a slightly increased t_2 .

Table 3 Overall classification accuracy for the simulated image (%)

Method	Overall accuracy	Kappa coefficient
PL+SVM	89.32 ± 0.34	0.8385 ± 0.0102
GL+SVM	86.83 ± 0.42	0.8011 ± 0.0154
α SVM	87.03 ± 0.36	0.8056 ± 0.0107
GL+LpcSVM with $n_k = 0$	89.08 ± 0.44	0.8346 ± 0.0125
GL+LpcSVM with $n_k \sim N(0, 0.05^2)$	88.97 ± 0.52	0.8331 ± 0.0121

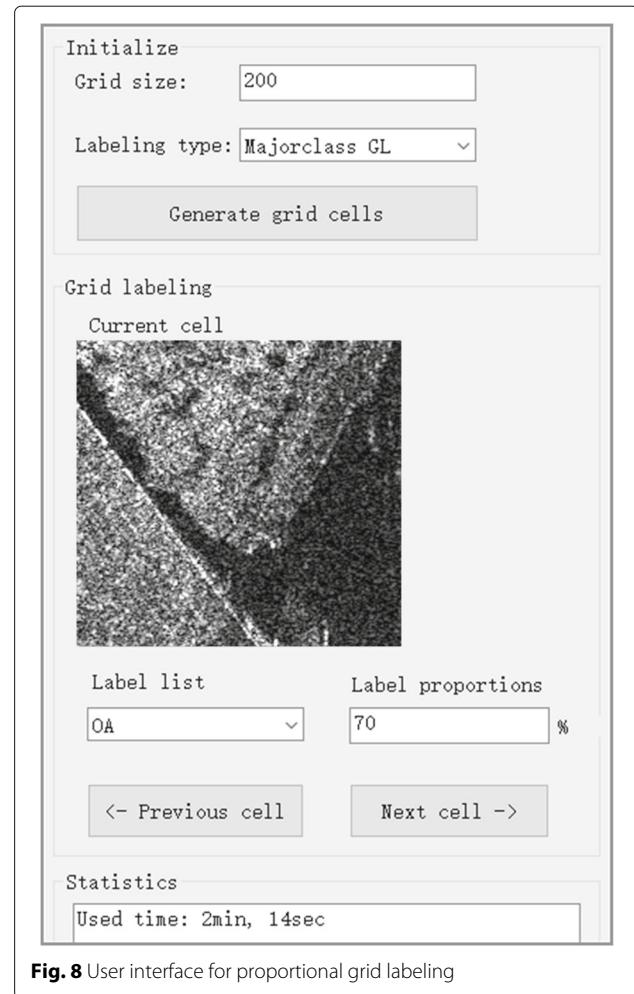


Fig. 8 User interface for proportional grid labeling

In addition, the cost of land cover classification has been reduced significantly with the reduction of training set definition time. More importantly, expensive field surveys can be avoided for sample labeling, making our method suitable for practical large-scale SAR image understanding.

6 Conclusions

This paper presents an efficient SAR image classification method through learning from label proportions. Grid labeling is introduced to obtain the truth of training data more efficiently. In order to eliminate label uncertainty coming from grid labeling, we present an SVM-based

Table 4 Average time used for classification of Tianjin and Rosenheim areas (min)

Method	Training set definition time t_1	Classifier consuming time t_2	Total time t
PL+SVM	514	22	536
GL+LpcSVM	128	37	165

model for learning from label proportions. A proximate inference algorithm is proposed through defining a reweighting strategy which has considered label agreement and label proportions. Our method not only outperforms α SVM, a state-of-the-art approach on learning from label proportions, but also demonstrates comparable accuracy with traditional classification methods working with pixel labeling, while bringing great reduction of labeling cost for training data. Future work includes developing an automatic estimation method for label proportions to further reduce human labor in sample labeling and extending our method to more application scenarios, such as SAR urban LULC mapping [3, 14] and multispectral image classification.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 61331015 and U1406404.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Center for Advanced Sensing Technology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China. ²School of Aeronautics & Astronautics, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China.

Received: 9 March 2016 Accepted: 15 May 2017

Published online: 31 May 2017

References

1. Y Ding, Y Li, W Yu, SAR image classification based on CRFs with integration of local label context and pairwise label compatibility. *IEEE J. Selected Topics Appl. Earth Observ. Remote Sensing*. **7**(1), 300–306 (2014). doi:10.1109/JSTARS.2013.2262038
2. JA Richards, X Jia, *Remote sensing digital image analysis: an introduction*. (Springer, Berlin, 2006)
3. P Gamba, M Aldrichi, SAR data classification of urban areas by means of segmentation techniques and ancillary optical data. *IEEE J. Selected Topics Appl. Earth Observ. Remote Sensing*. **5**(4), 1140–1148 (2012). doi:10.1109/JSTARS.2012.2195774
4. C Tison, F Tupin, H Maitre, A fusion scheme for joint retrieval of urban height map and classification from high-resolution interferometric SAR images. *IEEE Trans. Geosci. Remote Sensing*. **45**(2), 496–505 (2007). doi:10.1109/TGRS.2006.887006
5. S Rajan, J Ghosh, MM Crawford, Exploiting class hierarchies for knowledge transfer in hyperspectral data. *IEEE Trans. Geosci. Remote Sensing*. **44**(11), 3408–3417 (2006). doi:10.1109/TGRS.2006.878442
6. C Persello, L Bruzzone, Active learning for domain adaptation in the supervised classification of remote sensing images. *IEEE Trans. Geosci. Remote Sensing*. **50**(11), 4468–4483 (2012). doi:10.1109/TGRS.2012.2192740
7. C Persello, Interactive domain adaptation for the classification of remote sensing images using active learning. *IEEE Geosci. Remote Sensing Lett.* **10**(4), 736–740 (2013). doi:10.1109/LGRS.2012.2220516
8. B Demir, F Bovolo, L Bruzzone, Classification of time series of multispectral images with limited training data. *IEEE Trans. Image Process.* **22**(8), 3219–3233 (2013). doi:10.1109/TIP.2013.2259838
9. N Quadrianto, AJ Smola, TS Caetano, QV Le, Estimating labels from label proportions. *J. Mach. Learn. Res.* **10**, 2349–2374 (2009)
10. S Rueping, in *Proceedings of the 27th International Conference on Machine Learning*. SVM classifier estimation from group probabilities (ACM, Haifa, 2010), pp. 911–918
11. F Yu, D Liu, S Kumar, T Jebara, S Chang, in *Proceedings of the 30th International Conference on Machine Learning*. α SVM for learning with label proportions (ACM, Atlanta, 2013), pp. 504–512
12. G Patrini, R Nock, T Caetano, P Rivera, in *Advances in Neural Information Processing Systems 27*, ed. by Z Ghahramani, M Welling, C Cortes, ND Lawrence, and KQ Weinberger. No label no cry (Curran Associates, Inc., North Miami Beach, 2014), pp. 190–198
13. SS Keerthi, S Sundararajan, S Shevade, in *Proceeding of the 24th International Conference on Computational Linguistics*. Extension of TSVM to multi-class and hierarchical text classification problems with general losses (Association for Computational Linguistics, Mumbai, 2012), pp. 1091–1100
14. Y Ding, L Qiu, P Yang, Z Zhou, Y Li, W Yu, in *2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Scene scattering descriptor for urban classification in very high resolution SAR images, (IEEE, Melbourne, 2013), pp. 2015–2018. doi:10.1109/IGARSS.2013.6723205
15. C Cortes, V Vapnik, Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995). doi:10.1007/BF00994018
16. C-C Chang, C-J Lin, LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011). doi:10.1145/1961189.1961199
17. W Dai, Q Yang, G-R Xue, Y Yu, in *Proceedings of the 24th International Conference on Machine Learning*. Boosting for transfer learning (ACM, Corvallis, 2007), pp. 193–200. doi:10.1145/1273496.1273521
18. Y Freund, RE Schapire, A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**(1), 119–139 (1997). doi:10.1006/jcss.1997.1504
19. ZH Zhou, *Ensemble methods: foundations and algorithms*. (Taylor & Francis, New York, 2012)
20. G Jun, J Ghosh, in *2008 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis, vol. 1 (IEEE, Boston, 2008), pp. 52–55. doi:10.1109/IGARSS.2008.4778790
21. T-F Wu, C-J Lin, RC Weng, Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res.* **5**, 975–1005 (2004)
22. GH Hardy, JE Littlewood, G Pólya, *Inequalities*. *Cambridge Mathematical Library*. Section 10.2, Theorem 368. (Cambridge University Press, Cambridge, 1952)
23. D Dai, W Yang, H Sun, Multilevel local pattern histogram for SAR image classification. *IEEE Geosci. Remote Sensing Lett.* **8**(2), 225–229 (2011). doi:10.1109/LGRS.2010.2058997
24. JS Lee, E Pottier, *Polarimetric radar imaging: from basics to applications*. (CRC Press, Boca Raton, 2009)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com