






Discriminative Sketch Topic Model With Structural Constraint for SAR Image Classification

Yake Zhang, Fang Liu , Senior Member, IEEE, Licheng Jiao , Fellow, IEEE,
Shuyuan Yang , Senior Member, IEEE, Lingling Li , Member, IEEE,
and Meijuan Yang , Graduate Student Member, IEEE

Abstract—Synthetic aperture radar (SAR) image classification is an important part in the understanding and interpretation of SAR images. Each patch in SAR images has a scene category, but usually contains multiple land-cover classes or latent properties, which can be represented by topics in the probabilistic topic model (PTM). The representation and selection of discriminative features in PTM have a large impact on the classification results. Most of the existing feature learning methods do not make full use of high-level structure feature and the feature correlation within similar images to mine discriminative features. Therefore, this article proposes a discriminative sketch topic model with structural constraint (C-SSTM) for SAR image classification. In the proposed model, each image patch is characterized by structural and texture features. In particular, the sketch structural feature is based on the sketch map to represent the image local structure pattern. Then, the local image manifold information is preserved in terms of structure and texture. In the structural constraint, the texture and structure of each image patch are combined to learn discriminative latent semantic topics between image patches. Finally, each image patch is quantified by discriminative latent semantic topics instead of low-level representation. The experimental results tested on synthetic and real SAR images demonstrate that the proposed C-SSTM is able to learn effective structural feature representation from SAR images. Compared with other related approaches, C-SSTM produces competitive classification accuracies with high time efficiency.

Index Terms—Image classification, local image manifold, probabilistic topic model, sketch structural feature, synthetic aperture radar (SAR) image.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) systems are capable of working under various seasons and weather conditions

Manuscript received July 14, 2020; revised September 3, 2020; accepted September 8, 2020. Date of publication September 15, 2020; date of current version September 30, 2020. This work was supported in part by the State Key Program of National Natural Science of China under Grant 61836009, in part by the Key Research and Development Program in Shaanxi Province of China under Grant 2019ZDLGY03-06, in part by the Program for Cheung Kong Scholars, in part by the Innovative Research Team in University under Grant IRT_15R53, in part by the Fund for Foreign Scholars in University Research and Teaching Programs (the 111 Project) under Grant B07048, and in part by the National Natural Science Foundation of China under Grant 62076192. (Corresponding author: Fang Liu.)

The authors are with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, and the Joint International Research Laboratory of Intelligent Perception and Computation, International Research Center for Intelligent Perception and Computation, School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: yake_1023@163.com; f63liu@163.com; lchjiao@mail.xidian.edu.cn; syyang@xidian.edu.cn; linglingxidian@gmail.com; mjuanyang@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2020.3024002

[1], [2]. They have been widely used for applications including environmental surveillance and regional planning [3]. With the rapid development of remote sensing technologies, a large number of SAR images are now available. This situation makes manual interpretation a time-consuming and expensive process. The SAR image classification as an important part of image interpretation has attracted more and more attention [4].

The classification of SAR images depends on the representation and selection of discriminant features [5]. Recently, various feature extraction approaches have been developed to characterize the content of SAR images [6]–[8]. From the perspective of the human visual system, the content of an image usually consists of two parts, which represent the appearance and structure information [9]. To characterize the appearance of SAR images, many researchers utilized the spectral and texture descriptors, such as the gray-level co-occurrence matrix (GLCM) [10], [11]. Other efforts [12] have explored the wavelet transform to extract texture features, in which the kurtosis value of the wavelet energy feature is utilized to describe statistical information. With the increase of resolution, it is becoming more critical to model the complex structures in the local content. The spatial structures are employed to characterize the local shape information of SAR images, which play an important role to recognize and distinguish terrains. Many transform domain filters, for instance, Gabor transform filters [13], [14], wavelet transform filters [15], and curvelet transform filters [16], have been employed to extract spatial structures of terrains. However, it is difficult to decide the suitable parameters of scales and orientations of these transform filters, which need a large number of experiences and prior knowledge. The scale-invariant feature transform (SIFT) [17] and its variants [18] are the most popular approaches for local structure description in remote sensing images. Due to its efficiency, it is widely used to distinguish and recognize terrains for classification applications [19], [20]. First, the keypoint detection is carried out by the SIFT algorithm, which directly affects the subsequent orientation assignment and descriptors extraction. However, the keypoint detection is seriously affected by speckle noise in the SAR image. Although the statistical specificity of the SAR image is considered, there are still many false keypoints detected by literature [18]. In [21], a weighted neighborhood filter bank is proposed, which can extract spatial information and achieve a high discriminative power.

Deep learning-based methods [22]–[24] extract the features from the images in a joint spatial-spectral manner [25]. It has

turned out to be good at modeling the intricate structures hidden in high-resolution images for segmentation and classification tasks [26]. In [15], a convolutional-wavelet neural network is proposed to compute structural features that account for the neighborhood of an individual pixel. By designing a spatial feature learning network based on long–short-term memory [24], the spatial dependencies of the SAR image were extracted automatically. Recently, a distribution and structure match auxiliary classifier generative adversarial network (DSM-ACGAN) [27] was proposed to optimize structure features obtained by adversarial learning, which combined the characteristics of statistical distribution and spatial structure. Although these networks produce superior accuracies, they usually require a large number of training samples and parameters in the convergence of the network [28]. The above approaches are designed to extract structure descriptors based on pixel space, which suffer from enormous influence by speckle noise. In addition, methods of computing the structure descriptors only in pixel space lack the ability to capture the high-level semantic relationship.

For high-resolution SAR images, different remote sensing scenes are usually composed of distinct thematic classes. For example, an image patch associated with a scene representing industrial might contain some thematic categories such as trees, buildings, parking lots, and roads. The variability and ambiguity of different scenes make scene representation and recognition a challenging task [29]–[31]. Therefore, there is an increased interest and demand to enhance the discrimination of learned features and improve the classification accuracy.

In order to improve the discrimination of feature descriptors, a few studies on feature coding are proposed to reduce the “semantic gap” [32] between the low-level data and high-level semantic information. For instance, Hou *et al.* [33], introduced a hierarchical sparse representation classification to precisely describe the complex land-cover objects by exploiting the multisize patches around each pixel. Zhao *et al.* [34] proposed a semisupervised feature learning approach to extract the discriminative information which was derived from both the labeled and unlabeled SAR image patches in a sparse ensemble learning framework. These methods are based on dictionary learning, which yields dictionary atoms and sparse coefficients for the subsequent classification. The bag of words (BOW) [35], [36] model has also shown excellent representational capacity. Inspired by the BOW model, the probabilistic topic model (PTM) [37] represents the imagery as a random mixture of topics. The PTM includes probabilistic latent semantic analysis (PLSA) [38], latent Dirichlet allocation [39], and fully sparse topic model [40]. The PTM has been applied to natural scene interpretation in [41]. This model is suitable not only for the natural image but also for the remote sensing image. Yang *et al.* [42] combined the benefits of quadtree representations and aspect models to improve local label consistency and sharpen the labelings in a broader context. Zhu *et al.* [20] explored a sparse homogeneous and heterogeneous information from image blocks at the same time, and discovered discriminatory semantic descriptions by topic models. However, the above methods are all aimed at the optimization and improvement of multiple features of the same image patch; the relevance of topics among different

image patches is not considered. Recent studies have shown that the semantically similar features usually co-occur in similar images with a high probability [43]. Namely, analogous images in nearby manifold space often have close latent semantic topics. Using this property, we can further improve the robustness and discrimination of latent semantic topics.

Based on the above analysis, we propose the discriminative sketch topic model with structural constraint (C-SSTM). C-SSTM alternately explores the sketch map [44], [45] and pixel space to capture the sketch structural feature. The sketch map constituted by sketch line segments describes the change of pixel amplitude in SAR images, and it is a sparse representation for the image structure. Compared with structure features obtained in pixel space, the sketch structural feature includes high-level semantic information and is more robust to speckle noise. In addition, the structural constraint constructed by a nearest-neighbor graph in local image manifold is also introduced to the PTM for enhancing the discriminative capacity of learned features. The major contributions of this work are listed as follows.

- 1) A sketch map represents the sparse spatial relationship of pixels, which contains crucial structure information. We utilize the complementary information of the sketch map and pixel space to construct the sketch structural feature. Thus, the sketch structural feature is more robust to speckle noise and possesses more comprehensive structural information.
- 2) The image manifold consists of the nearest-neighbor graph, representing intrinsic structures of images. According to image manifold, both structure and texture descriptors are considered to structural constraint for getting more accurate latent semantic topics of PTM.
- 3) Experiments are conducted on both synthetic and real SAR images, indicating that the proposed C-SSTM achieves promising performances in terms of classification accuracy and time consumption.

The remainder of this article is arranged as follows. The related works on the sketch map and PTM are presented in Section II. In Section III, the whole framework of C-SSTM is presented. The sketch structural feature extraction and structural constraint in the C-SSTM model are described in detail. Experimental results and analysis are carried out in Section IV. Finally, conclusions are presented in Section V.

II. RELATED WORK

A. SAR Sketch Map

Considering the statistical distribution, speckle, and geometric characteristics of SAR images, Wu *et al.* [44] proposed the SAR image sketching model. For image structures, the sketch map is a sparse representation at a high semantic level. The main process includes designing the edge-line templates, extracting the curves, sketch lines approximation, and preserving the significant sketch lines.

The primitive of the sketch map is a sketch line; each sketch line consists of several sketch line segments which are connected end to end. The sketch line segments represent the boundary of two different objects in high-resolution SAR images, which are

the same as optical images. Meanwhile, the sketch line segments can represent a line target such as a bridge, or land-cover objects such as a house formed by some boundaries created by the bright spots and shadows. That is to say, in addition to representing the boundary of different objects in high-resolution SAR images, the sketch line segment can also sparsely represent objects higher than the ground. Therefore, the sketch line segment is a sparse representation of high-resolution SAR image structure in semantic space. The sketch line segment not only represents the change of gray value, but also obtains the detailed and sparse structure at the changed pixels. Namely, the sketch line segment describes the direction and position of the structure information in SAR images.

B. Probabilistic Topic Model

The PLSA model is contained in PTM, which introduces a latent variable to analyze the probability distribution of visual words. First, proposed in information retrieval by Hofmann [38], PLSA has been extended to image interpretation due to the similarity between natural language analysis and image processing [46]. PLSA is able to map the low-level features to high-level latent semantic representations.

Given a dataset of M images $D = \{d_1, d_2, \dots, d_M\}$, each image can be described by a visual dictionary $W = \{w_1, w_2, \dots, w_N\}$, and N is the number of visual words in the dictionary. The dataset can be represented as a word-image co-occurrence matrix; each element of the co-occurrence matrix is represented by $occ(d_m, w_n)$. $occ(d_m, w_n)$ denotes the number of times the visual word w_n occurred in image d_m . The latent topic set is $Z = \{z_1, z_2, \dots, z_T\}$, and T denotes the number of the topic. By choosing a topic z_t to the image d_m with probability $P(z_t|d_m)$, and a word w_n to the topic z_t with probability $P(w_n|z_t)$, the probability $P(w_n|d_m)$ can be decomposed as follows:

$$P(w_n|d_m) = \sum_{t=1}^T P(w_n|z_t)P(z_t|d_m). \quad (1)$$

The probability to select an image d_m can be defined as $P(d_m)$. Defining a log-likelihood function L

$$\begin{aligned} L &= \sum_{m=1}^M \sum_{n=1}^N n(d_m, w_n) \log P(d_m) P(w_n|d_m) \\ &= \sum_{m=1}^M n(d_m) \left[\log P(d_m) + \sum_{n=1}^N \frac{n(d_m, w_n)}{n(d_m)} \log P(w_n|d_m) \right] \\ &\propto \sum_{m=1}^M \sum_{n=1}^N n(d_m, w_n) \log \sum_{t=1}^T P(w_n|z_t) P(z_t|d_m). \quad (2) \end{aligned}$$

The goal of PLSA is to learn a model that gives a high probability to the word-image co-occurrence. The parameters $P(w_n|z_t)$ and $P(z_t|d_m)$ are obtained by maximizing the log-likelihood function L . Then, the image topic probability distribution $P(z_t|d_m)$ is the latent semantics that we intend to mine, and image $d_m (m = 1, 2, \dots, M)$ is represented by the vector $\{P(z_1|d_m), P(z_2|d_m), \dots, P(z_T|d_m)\}$.

III. METHOD OF C-SSTM MODEL

To effectively learn discriminative semantic topic representations, the C-SSTM model is proposed for SAR imagery classification. The overall flowchart of SAR classification based on the C-SSTM model is presented in Fig. 1. Our approach consists of three main steps. The first step is the structure and texture visual words generation [Fig. 1(a)]. We split the images into image patches using a uniform grid sampling strategy, and digitize the patches by corresponding texture and structure features. Then, a k -means method is applied to generate the texture and structure visual words, respectively. In the second step [Fig. 1(b)], the discriminative latent semantic is mined. Using local image manifold to discover intrinsic information of images is based on the fact that the semantically similar features generally co-occur in similar images with a high probability. The next step [Fig. 1(c)] is to feed the discriminative latent semantic into the support vector machine (SVM) classifier. Finally, we get the scene labels of each image patch.

A. Feature Extraction

According to the biological vision, Chang and Tsao [9] found that when a face image is represented by a vector containing shape and appearance information, each neuron has its own specific vector in the inferior temporal cortex region. When encoding facial images, the face cell's response is proportional to the projection of a face stimulus onto shape and appearance dimensions. Similar to human faces, ordinary objects also have shape and appearance information. So the structure and texture features are important for characterizing image patches, and we use these two complementary features to represent an image.

The sketch map is a sparse representation of the image structure, where the direction and length of the boundary can be defined by sketch line segments. The image structure contains not only boundary information, but also the interrelationship between the boundaries. At the same time, the pixel space contains more original information, so we define the sketch structural features based on the interaction between the sketch map and the pixel space. To better illustrate the sketch structural feature extraction procedure, we use a real SAR image block as an example. The SAR image and its corresponding sketch map are displayed in Fig. 2(a) and (b). The binary image [Fig. 2(c)] is obtained by Otsu's method, dividing the SAR image into foreground and background region. Let B denote the binary image, where the part of 1 in B corresponds to the foreground region, and the part of 0 in B corresponds to the background. Multiple land-cover objects are collected in the foreground region. The points in the foreground region are aggregated according to 4-connected rule, and several separate regions are obtained. These regions are sorted according to the size, and the main object in the SAR image is deemed to be the largest one. As shown in Fig. 2(d), two regions are marked in red and green, respectively. The size of the red region is larger than the green region, so the main object of the SAR image is the red portion. Depending on the sketch map, the distance transformation method finds the sketch point associated with each point in the SAR image. The sketch line segment where

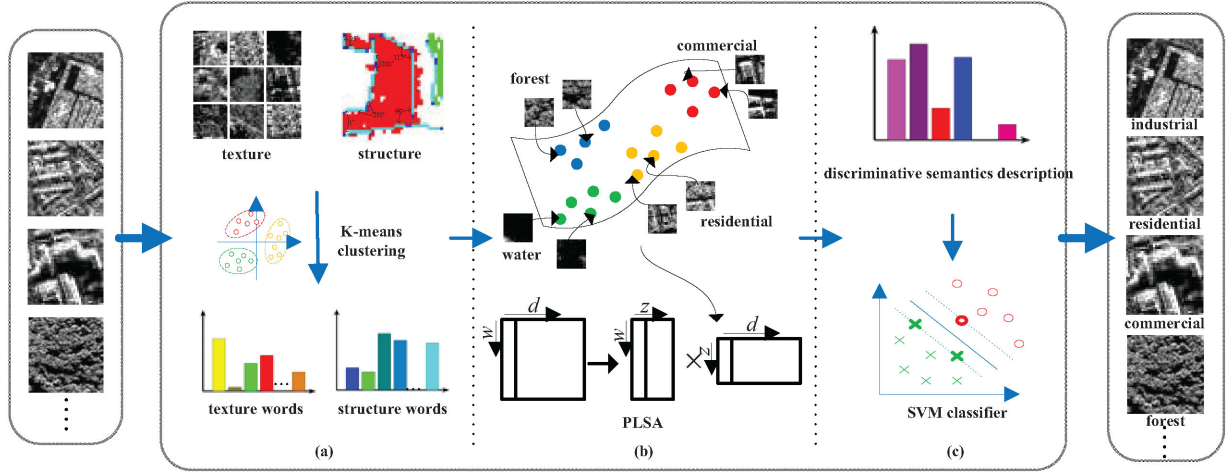


Fig. 1. Flowchart of SAR image classification framework based on the C-SSTM. The parameters w , d , and z represent the images, visual words, and topic variables, respectively.

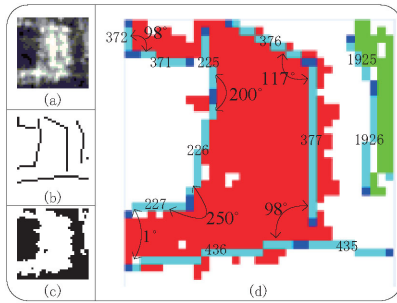


Fig. 2. A sketch structure feature example. (a) SAR image. (b) Sketch map. (c) Binary image. (d) Angle between sketch lines.

the sketch point is located has an index, which is the same index with the associated pixels. In Fig. 2(d), the sketch line segments are highlighted in light blue, the endpoints of each sketch line segment are shown in blue, and the value at the middle position of each line segment is the index number. The set of line segment indexes associated with the foreground region is the equivalent set associated with all points in this region. Here we define the point–line association criterion.

Let $w(i)$ denote a small patch with 5×5 pixel in size around pixel i , and the relevance degree $H(i)$ between pixel i and land-cover object is defined as follows:

$$H(i) = \sum_{n \in w(i)} B(n). \quad (3)$$

The starting, middle, and terminal pixel of the sketch line segment l are marked as l_s , l_m , and l_t , respectively. The indicator function $\delta(i)$ is

$$\delta(i) = \begin{cases} 1, & H(l_s) \geq 3 \text{ or } H(l_t) \geq 3 \\ 2, & H(l_m) \geq 5 \\ 0, & i \in l / \{l_s, l_m, l_t\}. \end{cases} \quad (4)$$

The relevance degree $F(l)$ between sketch line segment l and the red region is defined as

$$F(l) = \sum_{i \in l} \delta(i). \quad (5)$$

If $F(l)$ is greater than 3, the sketch line segment l is associated with the object, otherwise it is not associated.

Since the sketch line segment 435 in Fig. 2(d) does not satisfy the relevance degree $F(l)$, we obtain the sketch line segments associated with the red region, such as 372, 371, 225, 226, 227, 436, 377, and 376 in Fig. 2(d). The sketch line segment is connected according to the proximity of the end points of each line segment. In the set of line segments associated with the red region, Depth First Search algorithm searches the adjacency order, and starts with the line segment of the smallest index. In Fig. 2(d), the longer branch in the Depth First Search algorithm is a set of line segments with index 225, 226, 227, 436, 377, and 376. Then, the angle between two adjacent sketch line segments is calculated. Based on the direction of two sketch segments and their relationship to the red region, the angle θ between the two sketch segments changes from 0° to 360° as shown in Fig. 2(d). The length of the sketch line segment is represented by len . Based on the above analysis, we define the ternary feature, including the length of two sketch line segments and the relative angle. Such as for the sketch line segments 436 and 377, $len(436) = 25$, $len(377) = 23$, and $\theta = 98^\circ$. The ternary feature is $(25^\circ, 23^\circ, 98^\circ)$ for these two sketch segments. The set of ternary features is utilized to represent the sketch structural feature of the image. The sketch structural feature descriptor extraction based on the sketch line segments is implemented as in Algorithm 1.

The texture descriptor is described by GLCM in this article. The GLCM uses the spatial correlation of gray to describe the texture, which is a widely used texture statistical analysis method [47]. From the computational efficiency and the storage of co-occurrence matrices, the gray level of the original image is often compressed first, and then the GLCM is calculated. Haralick *et al.* [48] extract quadratic statistics based on GLCM

Algorithm 1: Structural Feature Descriptor Extraction.**Input:** SAR image patches**Output:** The set of ternary features

- 1: Extracting the sketch maps of SAR image patches;
- 2: Generating a binary image B by Otsu's method, and the SAR image is divided into foreground and background region;
- 3: The points in the foreground region are aggregated according to 4-connected rule, and several small regions are obtained. These regions are sorted according to the size, and the main object in the SAR image is considered to be the largest one;
- 4: According to the sketch map, the distance transformation method finds the sketch point associated with each point in the SAR image. The sketch line segment where the sketch point is located has an index, which is the index associated with the pixel point;
- 5: According to the relevance degree $F(l)$, the set of line segment indexes associated with the main object is the same set associated with all points in this region;
- 6: To search the set of line segments associated with the region, DFS algorithm is used. In order to get the adjacency order, DFS starts with the line segment of the smallest index;
- 7: Based on the direction of two sketch segments and their relationship to the foreground region, the angle θ between the two sketch segments is calculated. The ternary feature is defined by the length of two sketch line segments and the relative angle. The set of ternary features is utilized to represent the structural feature of the image.

to represent texture features. As it is done in [49], the texture features generated from GLCM are energy, contrast, correlation, entropy, and inertia.

B. PTM Based on Multiple Features and Structural Constraint

By obtaining the sketch structural feature and the GLCM texture feature, we concatenate these features to obtain the feature descriptor [50]. A PLSA model based on sketch structural features is established to learn potential topic distribution. The likelihood function of the PLSA model can be expressed as

$$P(d, w; \theta) = \prod_d \prod_w P(d, w)^{occ(d, w)} \quad (6)$$

where $occ(d, w)$ denotes the number of co-occurrence of image d and feature w , and $p(d, w)$ represents the joint probability density function. Introducing hidden variable topics $z = \{z_1, z_2, \dots, z_t, \dots, z_T\}$, the parameters of the PLSA model are the union of $p(w|z)$ and $p(z|d)$:

$$\begin{aligned} p(w|z) | (w, z) &\in \{1, 2, \dots, N\} \times \{1, 2, \dots, T\} \\ \cup p(z|d) | (z, d) &\in \{1, 2, \dots, T\} \times \{1, 2, \dots, M\}. \end{aligned} \quad (7)$$

The log likelihood function of the complete data (d, w, z) can be written as

$$L = \sum_{i=1}^M \sum_{j=1}^N occ(d_i, w_j) \log \sum_{t=1}^T p(d_i) p(z_t | d_i) p(w_j | z_t). \quad (8)$$

The traditional PLSA model treats each image separately and independently without considering image manifold information. However, if two images are similar, they usually locate in adjacent manifold structure with close latent semantic topics. Therefore, the structure and texture information of the overall image are also crucial for learning latent semantic relations.

Using manifold learning theory and graph theory, we construct a graph model with M vertices, where each vertex represents an image, and the weight of each edge denotes the similarity between two images. The higher the weight of the edge, the more similar the image. Conversely, when the weight of the edge is low, it means that the two images differ in structure and texture. Modeling the image manifold with the nearest neighbor graph provides additional information to learn latent semantic topics. The image correlation matrix $v(d_i, d_j)$ is defined as

$$v(d_i, d_j) = \exp(-Structure - Texture). \quad (9)$$

Let the *Structure* and *Texture* represent the similarity of structure and texture, respectively, and they are defined as follows:

$$Structure = 1 - \frac{SSIM(x_i, x_j)}{\sigma_1} \quad (10)$$

$$Texture = \frac{\|y_i - y_j\|_2^2}{\sigma_2} \quad (11)$$

where x_i is the sketch map of the image d_i , and the texture feature vector of image d_i is y_i . We use *SSIM* [51] to measure the structural similarity of sketch maps x_i and x_j , that is, the structural similarities of images d_i and d_j . σ_1 is the mean squared distance of all structural similarities *SSIM*(x_i, x_j), and σ_2 is the mean squared distance between two texture features y_i and y_j . When the sketch maps x_i and x_j are more similar, the *SSIM* will be larger and the structural item *Structure* will be smaller. When the texture features y_i and y_j are more similar, the texture item *Texture* will be smaller. When *Structure* and *Texture* are both small, $v(d_i, d_j)$ will be large, indicating that d_i and d_j are similar.

Summarized from the above functions (9)–(11), we get the matrix $v(d_i, d_j)$. To maintain that two images with similar features have a large probability to share close topics in image manifold, we minimize the following function:

$$R = \sum_{t=1}^T \sum_{i,j=1}^M (p(z_t | d_i) - p(z_t | d_j))^2 v(d_i, d_j). \quad (12)$$

The R tends to make similar images have the same topics. When the structure and texture information are all similar at the same time, $v(d_i, d_j)$ is large; then, minimizing R strongly enforces the conditional probabilities $p(z_t | d_i)$ and $p(z_t | d_j)$ to take a close value. However, when the difference of images d_i and d_j is large, $v(d_i, d_j)$ is small. Thus, the discrepancy between

conditional probabilities $p(z_t|d_i)$ and $p(z_t|d_j)$ is allowable in optimization. Function R makes full use of image-level structure and texture information to learn the probabilistic topic distribution by considering the image manifold.

By combining sketch structural feature and texture feature, a multifeature joint PLSA model is established, and structural constraint is added to the PLSA model to optimize the probability distribution of latent semantic topics. The structural constraint R is incorporated after the log likelihood function L , and the final integrated function is defined as

$$\begin{aligned} L' &= L - \gamma R \\ &= \sum_{i=1}^M \sum_{j=1}^N \text{occ}(d_i, w_j) \log \sum_{t=1}^T p(d_i) p(z_t|d_i) p(w_j|z_t) \\ &\quad - \gamma \sum_{t=1}^T \sum_{i,j=1}^M (p(z_t|d_i) - p(z_t|d_j))^2 v(d_i, d_j) \end{aligned} \quad (13)$$

where γ is the regularization parameter. By maximizing the function L' , we get the parameters $p(z_t|d_i)$ and $p(w_j|z_t)$. The vector $\{p(z_1|d_i), p(z_2|d_i), \dots, p(z_T|d_i), d_i \in D\}$ represents the latent semantic topics for image d_i . The proposed model is based on sketch structure and texture information to construct multifeature PLSA. At the same time, it considers the influence of local image manifold on latent topic distribution, and mines the latent topics of images to obtain more distinguishable semantic features.

For the C-SSTM model, we use the generalized EM (GEM) algorithm in [52]. The GEM solves the log-likelihood function with hidden variables and realizes the maximum likelihood estimation. The GEM includes two steps: 1) give the posterior probability of hidden variables under the condition of the current estimated parameters (E -step); 2) increase the expected log-likelihood function of complete data (d_i, w_j, z_t) , and use the posterior probability of hidden variables obtained by E step to estimate the new parameters (M -step).

In the E -step, assuming that $p(z_t|d_i)$ and $p(w_j|z_t)$ are known, the posterior probability of the hidden variable $p(z_t|d_i, w_j)$ is obtained according to the Bayesian formula:

$$p(z_t|d_i, w_j) = \frac{p(w_j|z_t)p(z_t|d_i)}{\sum_{t=1}^T p(w_j|z_t)p(z_t|d_i)}. \quad (14)$$

In the M -step, when the joint probability of complete data (d_i, w_j, z_t) is known, the GEM aims to increase the expected log-likelihood function L' with the parameter $p(z_t|d_i)$ and $p(w_j|z_t)$. The GEM in [52] first maximizes L to obtain the initial estimate parameters $p(z_t|d_i)$ and $p(w_j|z_t)$, and then updates the regularization item in the function L' using Newton–Raphson method, so that the expected log-likelihood function L' is continuously improved. The formula of updating parameter $p(z_t|d_i)$ by Newton–Raphson method is as follows:

$$p(z_t|d_i)_{k+1} = (1 - \lambda)p(z_t|d_i)_k + \lambda \frac{\sum_{j=1}^M v(d_i, d_j) p(z_t|d_j)_k}{\sum_{j=1}^M v(d_i, d_j)}. \quad (15)$$

Algorithm 2: Implement of the C-SSTM Model.

Input: SAR image patches

Output: SAR image classification result

- 1: The SAR image is sampled by uniform grid and the set of image patches is obtained, in which the label of the image patch is the central pixel's label, and the label of the image patch needs to be predicted when the central pixel is not marked;
 - 2: According to Section III-A, the sketch structure and texture features of each SAR image patch are calculated;
 - 3: The sketch structure and texture features are clustered respectively through the K-means method, and each SAR image is represented by the connection of two BoW histograms;
 - 4: Initialize the probability $p(w_j|z_t)$ and $p(z_t|d_i)$;
 - 5: Maximize the expected log-likelihood function L' , in the E step, and calculate the implicit variable $p(z_t|d_i, w_j)$ according to formula (14); in the M step, according to the GEM algorithm in [52], first, obtain the initial estimate of the parameter $p(z_t|d_i)$ and $p(w_j|z_t)$, then iteratively update the regularization term according to formula (15), so that the value of the log-likelihood function L' is continuously increased until convergence;
 - 6: Take $p(z_t|d_i)$ as the representation of the image patch d_i , and send it and the corresponding label to the LibSVM with linear kernel function, in order to get the SVM model;
 - 7: Predict the label of the unmarked image patch based on the $p(w_j|z_t)$ and SVM model obtained from the training dataset;
 - 8: In the original SAR image, the label of each pixel is determined according to the voting criterion.
-

The step parameter λ controls how smooth the topic distribution is. The range of parameter λ is $(0,1)$. When λ takes a large value, $p(z_t|d_i)_{k+1}$ is mainly determined by $p(z_t|d_j)_k$ ($j \in \{1, 2, \dots, M\}$), that is, the new topic distribution mainly comes from the average of the neighborhood topic distributions in the previous step. When λ takes a small value, $p(z_t|d_i)_{k+1}$ is mainly determined by $p(z_t|d_i)_k$, and the new topic distribution mainly comes from the topic distribution in the previous step. The parameter λ affects the convergence rate, and does not affect the classification performance [52]. We set the constant parameter the same as in [52].

After obtaining the discriminative feature representation of the probability topic model, it is fed into the LibSVM classifier, which uses a linear kernel function. The overall implementation of the C-SSTM model is listed in Algorithm 2.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we first present the dataset introduction and experimental settings in Sections IV-A and IV-B, respectively. Section IV-C presents the computational complexity and time

TABLE I
PRIOR KNOWLEDGE OF THE SAR IMAGES

Images	Size	Resolution	Look	Band	Modalities	Location	Sensor
Bridge_los	350 × 480	1m	2	X	High resolution spotlight mode	New mexico	Airborne SAR
Pentagon	460 × 340	1m	2	Ku	Stripmap	Washington D.C	Airborne SAR
Noerdlingernew	540 × 470	1m	4	X	High resolution spotlight mode	Swabian Jura	TerraSAR

consumption of the algorithms. In Section IV-D, the selection of four hyperparameters is discussed. To verify the efficiency of the proposed method, the experimental results and analysis are conducted on the synthetic and real SAR images in Sections IV-E and IV-F.

A. Dataset

To verify the effectiveness and generalization of the C-SSTM model, the SAR images with different bands and sensors are utilized in the experiments. The detailed information of the three real SAR images is reported in Table I. The first real SAR image is the Bridge_los, which shows the urban area of Los Lunas, New Mexico. It was obtained from a subset of an X-band, high-resolution spotlight mode SAR data, acquired by the airborne SAR platform. The pixel size of this image is 350 × 480. Four major land-cover types are contained in Bridge_los, which are building, trees, idle land, and roads. The Pentagon image was obtained from a subset of a Ku-band, strip map SAR data, acquired by the airborne SAR platform, with a size of 460 × 340. The Pentagon data covers the area of Washington, DC. There are four major land-cover categories in this image, which are water, residential areas, commercial areas, and trees. The TerraSAR X-band data of Swabian Jura is used as the third real image. The image is acquired by a high-resolution spotlight mode, on Jul. 1, 2007, 23:00 UTC. The size of Noerdlingernew is 540 × 470. This image contains three categories of land-cover, which are industrial, residential, and farmland areas. The number of looks of the bridge_los, pentagon, and noerdlingernew are 2, 2, and 4, respectively. The original images and corresponding optical images are shown in Figs. 13(a) and (b)–15(a) and (b). In addition, the sketch maps and ground-truth images are provided in Figs. 13(c) and (d) and 15(c) and (d) to describe the complex spatial information and class labels. Within the ground-truth images, each color denotes one class and the unknown regions are labeled by white. The unknown regions are ignored during the training and testing phase.

We produce three synthetic SAR images to evaluate the performance of our method. The synthetic SAR image SYN1 contains two classes: residential area and commercial area, as shown in Fig. 7(a). The synthetic SAR image SYN2 contains three classes: residential area, commercial area, and forest, as shown in Fig. 8(a). The synthetic SAR image SYN3 contains four classes: water, agriculture, residential area, and forest, as shown in Fig. 9(a). For the three synthetic SAR images, we provide the corresponding ground truths [Figs. 7(b)–Fig. 9(b)] and sketch maps [Figs. 7(c)–Fig. 9(c)], and the size of synthetic SAR images is 180 × 180 pixels. Syn1 is the first synthetic SAR image that comes from residential and commercial areas

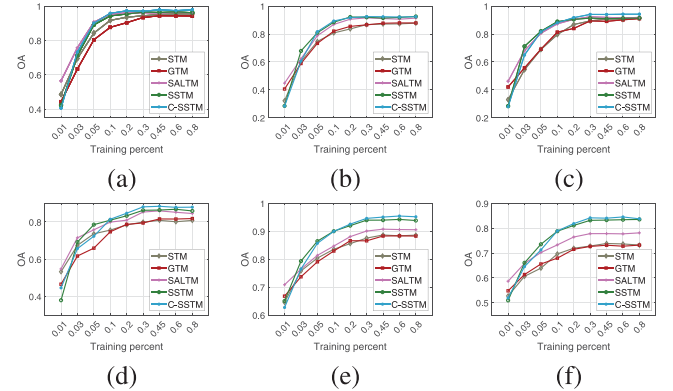


Fig. 3. OA of different training percent on the synthetic and real SAR images. (a) SYN1. (b) SYN2. (c) SYN3. (d) Bridge_los. (e) Pentagon. (f) Noerdlingernew.

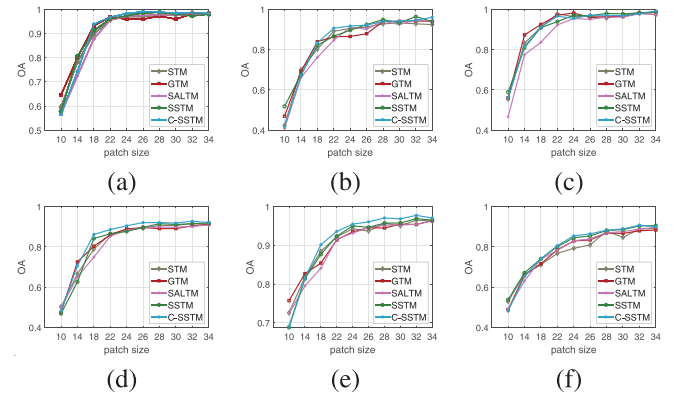


Fig. 4. OA of different patch sizes on the synthetic and real SAR images. (a) SYN1. (b) SYN2. (c) SYN3. (d) Bridge_los. (e) Pentagon. (f) Noerdlingernew.

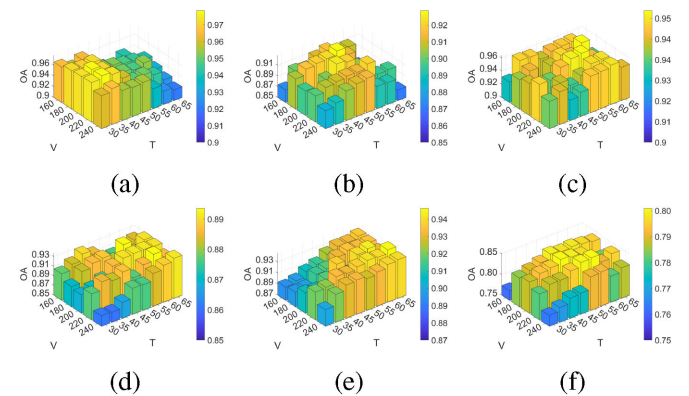


Fig. 5. OA versus different parameters V and T on the synthetic and real SAR images. (a) SYN1. (b) SYN2. (c) SYN3. (d) Bridge_los. (e) Pentagon. (f) Noerdlingernew.

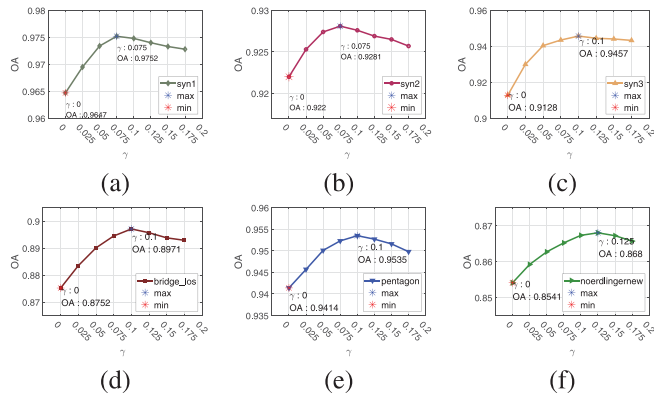


Fig. 6. OA of different γ on the synthetic and real SAR images. (a) SYN1. (b) SYN2. (c) SYN3. (d) Bridge_loss. (e) Pentagon. (f) Noerdlingernew.

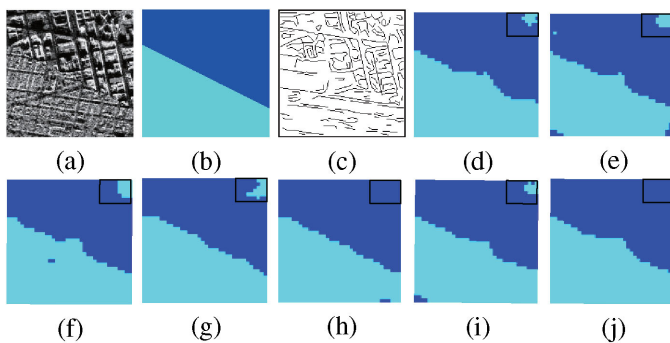


Fig. 7. Classification results of SYN1. (a) SYN1. (b) Ground truth. (c) Sketch map. (d) GTM. (e) STM. (f) SALT. (g) DCAE. (h) CWNN. (i) SST. (j) C-SSTM.

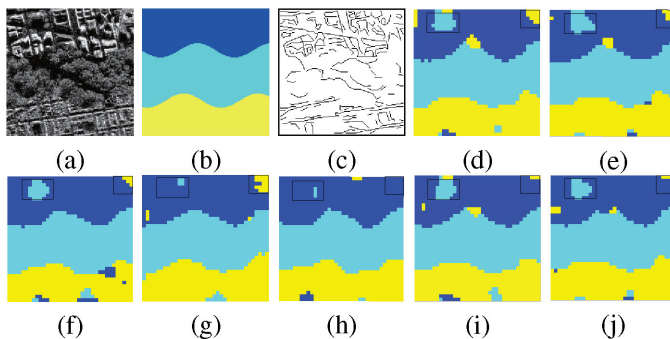


Fig. 8. Classification results of SYN2. (a) SYN2. (b) Ground truth. (c) Sketch map. (d) GTM. (e) STM. (f) SALT. (g) DCAE. (h) CWNN. (i) SST. (j) C-SSTM.

in the Pentagon. The composition of SYN2 is taken from the Pentagon’s commercial and residential areas, while the forest part is taken from the bridge_loss. The residential and water area in SYN3 are taken from the Pentagon, while the agriculture and forest area come from the Noerdlingernew and bridge_loss image, respectively.

B. Experimental Setup

In order to evaluate the efficiency of our proposed sketch structural features and structural constraint, we performed

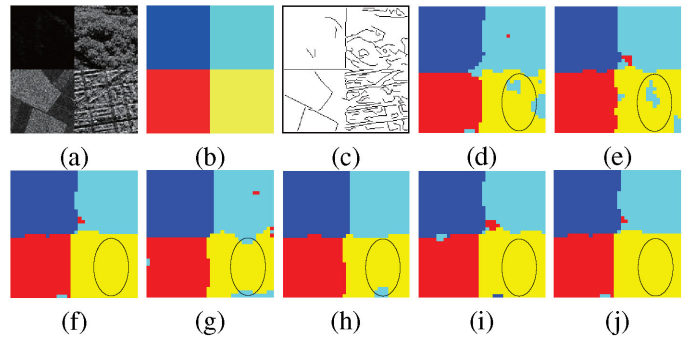


Fig. 9. Classification results of SYN3. (a) SYN3. (b) Ground truth. (c) Sketch map. (d) GTM. (e) STM. (f) SALT. (g) DCAE. (h) CWNN. (i) SST. (j) C-SSTM.

experiments on both synthetic and real SAR images. All the images are of 1-m resolution in experiments. In the feature extraction, we mainly use the structure and texture features of the image. The structural feature includes more macroscopic features of the image, and the texture features are more microscopic characteristics. The structural feature mainly includes shape information and the texture feature describes appearance information. This article focuses on the improvement of sketch structural feature and the structural constraint, and thus we use the same texture extraction method in related experiments.

To investigate the classification performance, several comparative experiments are designed as follows:

- 1) GTM: Gabor filter bank is used to extract the structure feature, and the structural response of SAR image is obtained by setting the direction and scale of Gabor filter; the mean and standard deviation of the Gabor coefficients are used to represent SAR image structure [13]. Then, the texture features obtained by the GLCM are combined and sent to the PLSA model to obtain a discriminative feature representation.
- 2) STM: The SIFT feature [18] has been widely applied in image analysis; it extracts the gradient direction histogram at the key points of the image as the structure features of the SAR image. The other settings in the experiment are the same as GTM.
- 3) SALT: A semantic allocation-level multifeature fusion strategy based on PLSA model [31] is employed to remote sensing image classification, in which the structure features use SIFT.
- 4) DCAE: DCAE [53] presented a deep convolutional autoencoders to learn discriminative features. In the DCAE method, a KL-divergence sparsity constraint is explored in layer-wise pretraining to model high-level features from original patches. For comparison with other methods, the DCAE network is trained without fine-tuning, connected with a linear SVM to yield classification results.
- 5) CWNN: The CWNN model [15] replaces the conventional pooling with a wavelet constrained pooling layer, keeping the structures of the learned features and suppressing the noise.

TABLE II
NUMBER OF LABELED SAMPLES IN EACH CATEGORY

Images	Total samples	Class 1	Class 2	Class 3	Class 4
SYN1	1024	512	512	-	-
SYN2	1024	341	341	342	-
SYN3	1024	256	256	256	256
Bridge_los	5930	200	200	200	200
Pentagon	5480	309	309	309	309
Noerdingernew	9160	1049	1049	1049	-

- 6) SSTM: The sketch structural features are based on sketch line segments, and the other experimental settings are the same as GTM.
- 7) C-SSTM: Different from SSTM, this model adds structural constraint to log likelihood function in order to get more discriminative latent semantic topics.
- 8) After the discriminative feature representation is obtained from the above model, it is fed into the classifier. Here we use the SVM classifier with the linear kernel function. For each category, we randomly select 50% samples of the labeled images, and others as the testing set. In order to obtain a convincing result, each method conducts 20 Monte–Carlo runs. More information about the labeled samples is provided in Table II, in which “–” represents that the information is not available. The performance is measured by the overall recognition accuracy (OA) and Kappa coefficient. OA is defined as the ratio of correctly classified pixels number to the total number; the Kappa coefficient is obtained from the confusion matrix to measure the consistency of the classification results.

In the experiment, we adopt a uniform grid sampling method. The patch size P is optimally set to 24×24 pixels with overlapping 5 pixels. In feature extraction, the structure and texture features performed well when the patch size and overlap are set to 8×8 pixels and 4 pixels [54], respectively. The number of visual words in the probability topic model is represented by V , and the number of topics is denoted by T . The setting of several vital hyperparameters is discussed in Section IV-D.

C. Computational Complexity and Time Consumption

The computation complexity of our method mainly exists in the procedure of feature extraction and topic model iteration. The feature extraction includes obtaining the sketch structural feature and texture feature. M samples are utilized in the experiment. Suppose that the average number of objects is N_o in each sample, and the average number of sketch lines is N_l around each object. When extracting texture features, there are N_g small patches of the sample. Therefore, the computation complexity of feature extraction is $O(M(N_o N_l + N_g))$. The topic model iteration includes the EM algorithm, the maximal iterative step is N_{iter} , and the computation complexity of E-step and M-step are $O(MNT)$ and $O(TN + MT + MN)$, respectively. So the complexity of the topic model is $O(N_{iter}(MNT + TN + MT + MN))$. In summary, the computation complexity of the proposed algorithm is $O(M(N_o N_l + N_g) + N_{iter}(MNT + TN + MT + MN))$. The experimental configuration of each method on the images is described in Sections IV-E and IV-F. All experiments

TABLE III
TIME CONSUMPTION OF DIFFERENT METHODS ON DIFFERENT IMAGES (S)

Image		GTM	STM	SALTM	DCAE	CWNN	SSTM	C-SSTM
SYN1	feature	194.41	2301.83	1998.65	180.36	-	76.05	78.64
	total	252.58	2275.80	2055.15	237.95	3922.13	148.08	206.59
SYN2	feature	152.48	2182.78	2071.81	164.74	-	74.30	73.31
	total	221.24	2268.41	2147.86	233.26	5548.42	139.16	187.87
SYN3	feature	160.37	2202.61	2022.58	167.03	-	70.22	76.05
	total	240.15	2290.86	2099.97	234.66	5889.86	145.84	196.67
Bridge_los	feature	319.41	4268.72	4204.65	329.05	-	154.84	144.77
	total	446.13	4439.88	4369.14	448.48	4155.42	262.98	348.53
Pentagon	feature	292.17	4322.96	4319.03	329.51	-	159.06	169.01
	total	438.64	4521.01	4506.14	470.44	6219.52	288.02	374.91
Noerdingernew	feature	680.24	10772.09	10574.21	746.64	-	424.08	433.38
	total	1084.30	11350.72	11152.47	919.79	18645.47	762.245	977.65

are implemented in MATLAB2018, with Intel Core I7 3.6-GHz CPU and 64-GB memory. The running time of each algorithm is reported in Table III, where the computational time of the feature extraction and total procedure are listed. The minimum time consumptions are highlighted by bold entities. The time consumption of feature extraction is not applicable in the CWNN algorithm. Since the feature extraction and classification are performed simultaneously, the labels are utilized to fine-tune the entire network. The feature extraction process of other methods can characterize image information without tags. Taken together, these results suggest our method achieves preferable computational efficiency than other algorithms.

D. Sensitivity Analysis

There are several parameters in the proposed C-SSTM, which are the training percent, patch size P , visual word number V , number of topics T , and regularization parameters γ . All the synthetic and real SAR images are utilized for the sensitivity analysis.

1) *Training Sample Percent*: To evaluate the sensitivity of different PLSAs with the number of training samples, Fig. 3 shows the classification accuracies with varying percent of training samples. The optimal parameters of the patch size, the visual word number V , the topic number T , and the regularization parameter γ are kept constant in the experiments. In each category, the random selection of the labeled samples is conducted over the range from 1% to 80%, and the rest as the testing set. To avoid bias induced by random sampling, we average the experimental results in the 20 Monte–Carlo runs [8].

As shown in Fig. 3, SSTM and C-SSTM obtain the unfavorable performance when the number of training samples is small. With the increasing of training percent, SSTM and C-SSTM outperform other methods achieving better classification accuracies. The reason is that SSTM and C-SSTM capture enough discriminative information with fewer feature dimensions. However, the other feature learning approaches capture more redundant information within features, leading to increased feature ambiguity and decreased class separability. In addition, the curves of accuracies in the SSTM and C-SSTM methods tend to be flat when the training percent is in the range of 20%–45%. The classification accuracies of other approaches increase slightly when the training percent reaches 45%. It is illustrated that SSTM and C-SSTM can provide optimal accuracies with fewer training percent. Taken together, these results suggest that our proposed methods have excellent representational capacity to

introduce enough discriminative information for image classification.

2) *Size of Patches*: An appropriate patch size can not only capture the enough contextual information, but also decrease the computational complexity. The influence of the patch size is discussed through the classification accuracy based on different PLSA-based methods for all synthetic and real SAR images. In these experiments, V , T , and γ were kept at the optimal parameter settings, respectively. The patch size is varied from 10 to 34. Fig. 4 shows the OA curves of the five comparative methods on synthetic and real SAR images.

From the perspective of the patch size, we can draw the following conclusions. With the increase of the patch size, the OA curves improve significantly within the first three sizes and achieve stable performance by increasing the patch size. We can see that after the patch size of 24×24 , the accuracy increases slowly. This indicates that when the patch size is 24×24 , it is enough to capture the local contextual information of the image. The larger size will increase the computational complexity, so we select the patch size to be 24×24 .

3) *Number of Visual Words V and Topics T* : The effect of the visual word number V and topic number T on the proposed C-SSTM is discussed in Fig. 5. The values of the patch size and γ were kept constant at 24 and 0.1, respectively. The visual word number V was then varied in the range of [160, 180, 200, 220, 240]. At the same time, the topic number T was varied from 30 to 65 with a stride of 5.

As can be seen in Fig. 5, when the value of V is fixed, the optimal topic numbers of different images are distinct. The optimal T values for C-SSTM with the six images are set to 35, 45, 50, 55, 60, and 50, respectively. Compared with three synthetic SAR images, the real SAR images have higher optimal T values. Moreover, with the increased complexity of the scene, the optimal T values are growing larger both in synthetic and real SAR images. These experimental results indicate that the optimal value of T depends on the scene complexity, since numerous terrain classes contained in the complex scene require more latent topics. When the value of T is fixed, the appropriate values of V vary from 180 to 200. The real SAR images require higher V values than synthetic images, and the reason is that the number of similar features in the large images is increased. To increase the classification accuracy, V and T should complement each other to robustly characterize SAR images.

4) *Regularization Parameter γ* : Finally, we check the classification performance of C-SSTM in relation to the regularization parameter γ on all images. The patch size P , the topic number T , and the visual word number V are kept constant. The regularization parameter γ is then varied from 0 to 0.175 with a stride of 0.025 in Fig. 6. The parameter γ controls the proportion of the original PLSA model and the structural constraint in exploring latent topics. When γ is 0, it is equivalent to performing model update based only on the original PLSA model. When γ is large, the structural constraint has a greater influence. Fig. 6 shows the variation of OA with the value of the regularization parameter γ . The optimal values γ vary from 0.075 to 0.1 on all the images. Experimental results validate the effectiveness of structural constraints in discovering discriminative latent topics.

TABLE IV
OPTIMAL k VALUES FOR THE DIFFERENT METHODS WITH THE THREE SYNTHETIC SAR IMAGES

Images	Methods				
	GTM	STM	SALTM	SSTM	C-SSTM
SYN1	35	35	35	30	30
SYN2	50	50	45	50	45
SYN3	60	60	60	60	50

E. Classification Results of the Synthetic SAR Images

For three synthetic SAR images, texture and structure features have the same number of dictionary $V/2$, where V is set to 200. The setting of topic number T of five PTM-based models refers to Table IV. To perform three synthetic images classification with the deep learning methods, the parameters of DCAE and CWNN follow the settings of methods [53] and [15], respectively.

Fig. 7 shows the classification results for SYN1, Fig. 8 shows the classification results for SYN2, and Fig. 9 shows the classification results for SYN3. Compared with the other PTM-based methods, the classification results using Gabor feature [Figs. 7(d)–9(d)] and SIFT feature [Figs. 7(e)–9(e)] are more heterogeneous. The Gabor feature is prone to confusion in places where texture is similar, especially at the boundary of two similarly textured objects. The result using the SIFT feature is better than that using the Gabor feature in most cases, especially for objects with significant structural differences. From the results of SALTM [shown in Figs. 7(f)–9(f)], we can see that changing the fusion strategy of multiple features can improve the classification performance. The structure feature of the SSTM is based on sketch map, and the sketch structure is obtained by sketch line segments, so it is robust to noise and surrounding objects. At the same time, the sketch structural feature also contains distinguishable structural information such as the size of the object. Therefore, the SSTM is superior to the GTM and STM methods in the classification performance of complex structures. The results of SSTM on SYN1 and SYN2 are better than that of SALTM. But on SYN3, the result by SSTM is not as good as that by SALTM. The reason is that the water and farmland in SYN3 contain less structural information, and the feature fusion strategy in SALTM makes full use of texture information. The classification results of C-SSTM [Figs. 7(j)–9(j)] are better in the consistency at the residential areas and forest.

For evaluating the classification performance in the complex structure areas, we highlight the classification results. The black rectangle in Fig. 7 highlights the notable classification results where the commercial is misclassified to residential. The results of GTM, STM, and SALTM are not satisfied in the commercial areas. The structure features of the above methods are greatly subject to the influence of speckle noise. Through calculating the original pixels, the Gabor feature is obtained by the mean and variance of the amplitude of the Gabor coefficient. The SIFT feature is a statistical representation of the direction histogram at the key point; it suffers from the same problem as the Gabor feature does. Nevertheless, the proposed sketch structural feature is based on the sketch map, and the sketch structural feature is a reasonable combination of sketch line

segments. Consequently, sketch structural feature is robust to speckle noise and has excellent discriminated capacity to distinguish the surrounding objects. At the same time, the sketch structural feature also contains the information of the size of objects, which is crucial for structural representation. It can be observed that the C-SSTM has significantly outperformed the SSTM method. This is due to the fact that C-SSTM utilizes the image manifold knowledge that similar images have similar latent topics to mine discriminative feature representations for subsequent image classification. Compared with deep learning methods, C-SSTM shows better visual effects than DCAE, and the classification results are comparable to CWNN. However, the computational requirements of deep learning methods are higher than that of C-SSTM, whereas the running time of C-SSTM is much shorter. In summary, the sketch structural feature captures enough discriminative information for image classification, and the proposed model is a highly efficient method with a low computational expend.

In Fig. 8, the remarkable misclassification results of commercial areas are highlighted by the black rectangle. It can be observed that the results of GTM, STM, SALT M, and SSTM are unsatisfactory. This can be explained by the fact that the commercial area in the black rectangle contains less structural information, where the superiority of the proposed sketch structural feature cannot be shown. The C-SSTM performs worse than the deep learning methods in the black rectangle. This is because the texture extraction method utilized by C-SSTM is relatively simple. In the black square, the commercial pixels are misclassified to residential class, and C-SSTM performs better than the GTM, STM, and SALT M methods. The classification results of C-SSTM in complex structural regions are comparable to those of CWNN.

As shown in Fig. 9, the water, agriculture, and forest areas of all methods are labeled perfectly because of its simple spatial structure. In the black ellipse, SALT M, SSTM, and C-SSTM yield satisfactory results. Since the structure in the residential region is not complicated, SALT M changes the fusion strategy of multiple features improving the classification performance. For deep learning methods, there has a misclassification phenomenon. Artificial buildings and trees are confused because the shadows caused by buildings are misclassified by high trees. However, C-SSTM has an excellent representation capacity of local structural information and discovers the feature correlation in image manifold to improve the consistency of regional classification.

From the classification results of synthetic SAR images, it can be illustrated that our model can obtain superior classification performance, especially in areas with complex structures. The sketch structural features are robust to speckle noise, and have excellent discriminant capacity in describing the complex image information. The structural constraint increases the feature separability in the manifold domain. Compared to deep learning methods, the proposed method makes full use of the image prior information, at a small-time computation, to obtain satisfactory performance. At the same time, the features extracted by deep learning are more general to all structure and texture features, and it is not special knowledge to structure or edge-line information.

TABLE V
OA(%) AND KAPPA COEFFICIENTS COMPARISON WITH THE THREE
SYNTHETIC SAR IMAGES

Methods	SYN1		SYN2		SYN3	
	OA	Kappa	OA	Kappa	OA	Kappa
GTM	94.40±0.23	0.8878±0.0011	87.88±0.28	0.8182±0.0019	88.99±0.32	0.8564±0.0017
STM	95.45±0.29	0.9091±0.0041	87.24±0.36	0.8034±0.0026	90.82±0.44	0.8775±0.0027
SALTM	96.36±0.21	0.9272±0.0016	91.40±0.32	0.8698±0.0024	92.00±0.37	0.9033±0.0028
DCAE	96.24±0.17	0.9248±0.0033	91.84±0.27	0.8902±0.0027	90.45±0.38	0.8727±0.0026
CWNN	98.07±0.15	0.9613±0.0028	93.57±0.24	0.9018±0.0021	93.02±0.36	0.9069±0.0031
SSTM	96.50±0.13	0.9299±0.0023	91.83±0.18	0.8761±0.0015	91.28±0.29	0.8836±0.0027
C-SSTM	97.52±0.10	0.9504±0.0016	92.81±0.20	0.8951±0.0019	94.57±0.22	0.9342±0.0018

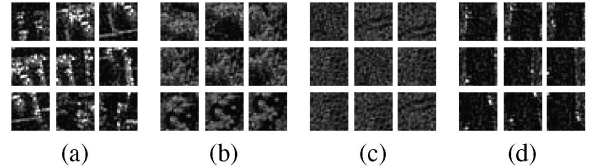


Fig. 10. Some labeled patches in Bridge_los. (a) Industrial. (b) Forest. (c) Idel land. (d) Road.

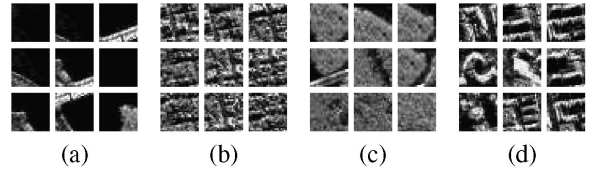


Fig. 11. Some labeled patches in Pentagon. (a) Water. (b) Residential. (c) Forest. (d) Commercial.

The numerical results are shown in Table V. In Tables V, VII, VIII, and IX, bold entities indicate the maximum values of OA and Kappa on each image. The OA and Kappa coefficient of STM on SYN1 and SYN3 are higher than GTM. The classification result of GTM on SYN2 is slightly better than STM. At the same time, the OA and Kappa coefficient of the GTM and STM are lower than the other three PTM-based methods. SSTM and SALT M have similar performance on numerical indicators, and SSTM's results are better on SYN1 and SYN2. C-SSTM is superior to other PTM methods in terms of the OA and Kappa coefficient. Compared with DCAE, C-SSTM obtains higher classification accuracy at a lower computational requirement. The classification results of C-SSTM are comparable to those of CWNN, especially in structurally complex areas, the C-SSTM achieves a higher classification accuracy. On the other hand, C-SSTM performs preferable computational capacity. From the visual classification results and numerical indicators, we can see the superiority of our method in structural feature extraction and learning discriminative latent topic features.

F. Classification Results of the Real SAR Images

In this section, three real SAR images are used to evaluate the classification performance. The relevant details of the three SAR images are presented in Table I. Some labeled patches of the the three real SAR images are shown in Figs. 10 –12. The majority vote method is used for the label of the overlapping portion of the image patches.

In the experiment, the visual word number V for different methods with Bridge_los, Pentagon, and Noerdlingernew are optimally set to 180, 200, and 220, respectively. The optimal

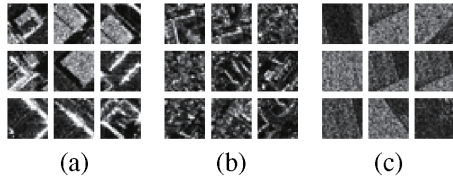


Fig. 12. Some labeled patches in Noerdlingernew. (a) Industrial. (b) Residential. (c) Farmland.

TABLE VI
OPTIMAL k VALUES FOR THE DIFFERENT METHODS WITH THREE SAR IMAGES

Images	Methods				
	GTM	STM	SALTM	SSTM	C-SSTM
Bridge_Jos	60	60	65	60	55
Pentagon	60	60	60	55	60
Noerdlingernew	55	55	50	45	50

TABLE VII
OA (%) AND KAPPA FOR BRIDGE_LOS

Methods	industrial	forest	idle land	road	OA	Kappa
GTM	87.37±0.81	85.30±0.49	94.14±0.30	60.00±1.59	81.65±0.34	0.7553±0.0055
STM	79.30±1.02	85.02±1.10	93.43±0.14	62.16±1.97	80.00±0.78	0.7334±0.0028
SALTM	91.58±1.24	81.69±0.43	94.14±0.41	74.05±1.39	85.39±0.43	0.8052±0.0053
DCAE	78.89±1.24	77.01±0.99	96.96±0.36	98.30±1.05	87.78±0.38	0.8308±0.0045
CWNN	79.95±0.95	80.56±1.07	97.91±0.54	99.38±0.73	89.45±0.65	0.8616±0.0072
SSTM	89.22±0.91	82.08±0.43	95.00±0.73	81.98±1.05	87.12±0.51	0.8402±0.0025
C-SSTM	94.65±0.83	85.02±0.71	94.62±0.30	83.03±1.51	89.73±0.33	0.8701±0.0028

TABLE VIII
OA (%) AND KAPPA FOR PENTAGON

Methods	water	residential	tree	commercial	OA	Kappa
GTM	99.22±0.15	62.27±0.14	96.74±0.24	94.89±0.23	88.27±0.16	0.8437±0.0029
STM	99.22±0.30	62.47±0.18	97.07±0.37	95.12±0.81	88.52±0.59	0.8469±0.0031
SALTM	99.22±0.35	70.52±0.27	96.74±0.45	94.16±0.79	90.15±0.25	0.8686±0.0037
DCAE	99.41±0.71	87.58±0.51	97.72±0.20	87.74±0.91	93.11±0.41	0.9081±0.0060
CWNN	99.84±0.36	89.91±0.26	99.47±0.23	91.43±0.49	95.16±0.25	0.9313±0.0022
SSTM	99.41±0.47	82.89±0.19	99.35±0.27	95.12±0.53	94.14±0.16	0.9218±0.0024
C-SSTM	99.61±0.75	87.01±0.44	98.70±0.20	95.89±0.77	95.35±0.22	0.9381±0.0023

TABLE IX
OA(%) AND KAPPA FOR NOERDLINGERNEW

Methods	industrial	residential	agricultural	OA	Kappa
GTM	78.82±0.54	50.37±0.55	90.24±0.51	73.14±0.25	0.5971±0.0031
STM	80.96±0.46	48.34±1.04	92.19±0.56	73.84±0.43	0.6076±0.0083
SALTM	75.68±0.55	66.76±0.65	90.95±0.86	77.78±0.66	0.6668±0.0069
DCAE	68.05±0.56	89.69±0.49	92.39±0.90	83.37±0.37	0.7498±0.0072
CWNN	82.77±0.63	89.69±0.78	90.73±0.83	87.74±0.64	0.8601±0.0066
SSTM	81.25±0.61	77.63±0.42	92.86±1.15	83.11±0.54	0.8066±0.0079
C-SSTM	84.71±0.42	80.54±0.68	94.63±0.96	86.81±0.62	0.8534±0.0046

numbers of topics T are shown in Table VI. The regularization parameter of the C-SSTM model is set to 0.1. The weight of the sparsity penalty in DCAE is 0.15 for three real SAR images, and the other hyper-parameters are the same as [53]. The CWNN method follows the parameter setting in [15].

From the visual results in Figs. 13–15, it is noted that in the complex land-cover scene, the classification results of GTM and STM appear more heterogeneous than the other three methods, particularly the road areas in Fig. 13, and the residential areas in Figs. 14 and 15. However, the SALTM focuses on the feature fusion strategy, and does not consider the feature extraction improvement and image manifold prior for extracting discriminant semantic features. Therefore, there are misclassification within residential, commercial, and industrial areas in Figs. 14(f) and 15(f). The SSTM does not use priori knowledge of similar latent topics within the similar images, resulting in unsatisfactory classification results, such as the industrial areas in Fig. 13(g),

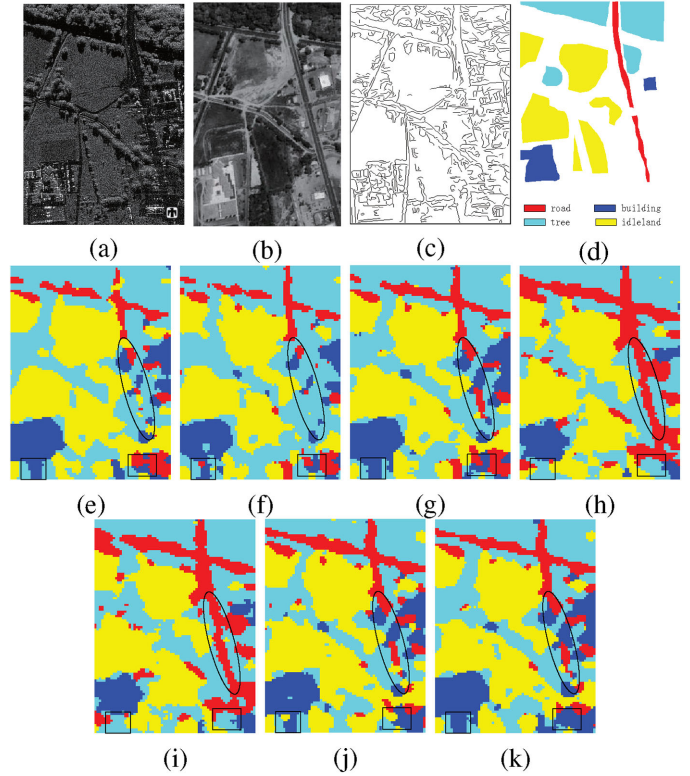


Fig. 13. Classification results of bridge_jos. (a) Bridge_jos. (b) Optical image. (c) Sketch map. (d) Ground truth. (e) GTM. (f) STM. (g) SALTM. (h) DCAE. (i) CWNN. (j) SSTM. (k) C-SSTM.

the commercial areas and residential areas in Fig. 14(g), and the residential areas in Fig. 15(g). Figs. 13(h)–15(h) show the classification results of the C-SSTM. We can see that in the complex land-cover scene such as residential areas, the classification consistency has been significantly improved compared with the other methods. At the same time, the classification accuracy in the water, farmland, and other homogeneous areas has also been improved. The reason is that we use the sketch structure based on sketch line segments and image manifold prior to extract discriminative latent topic representation. In Fig. 13, the shadow caused by trees is misclassified as road. To solve this problem, the sampling window needs to be increased, but this will reduce the efficiency of the method. Meanwhile, some idle lands with sparse trees are misclassified as forest. The water areas of these five methods are labeled perfectly because of its simple spatial structure. Compared with the classification results of all methods, it can be observed that the proposed C-SSTM can discover discriminative latent topics for images and show more robust and effective classification performance.

From the visual results in Fig. 13, it is observed that the classification results of GTM and STM are more heterogeneous than those of other PLSA-based methods, particularly in the ellipse area in Fig. 13. This is because the structure features extracted by the Gabor filters and the SIFT are affected by the speckle noise and surrounding objects, and it lacks enough discriminant structural information to characterize the different terrain types. Within the black rectangular, SSTM and C-SSTM perform better

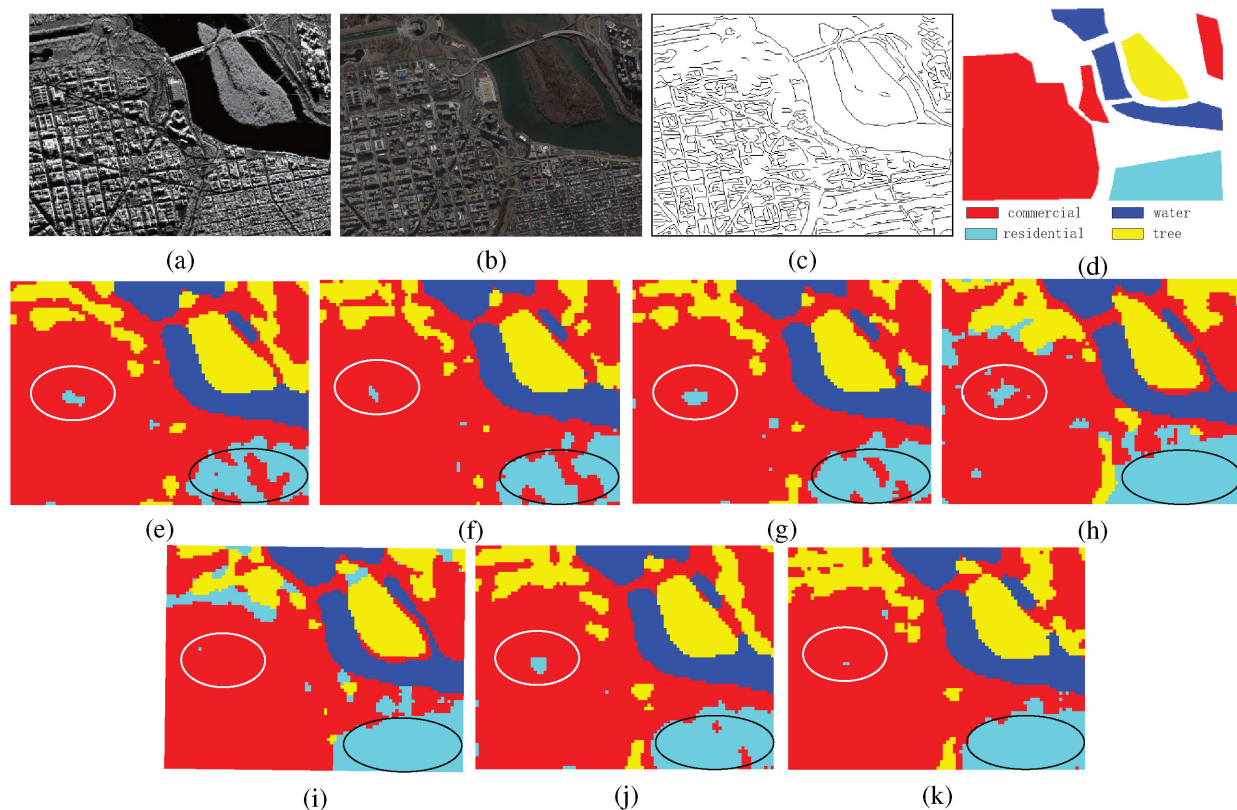


Fig. 14. Classification results of Pentagon. (a) Pentagon. (b) Optical image. (c) Sketch map. (d) Ground truth. (e) GTM. (f) STM. (g) SALT M. (h) DCAE. (i) CWNN. (j) SSTM. (k) C-SSTM.

than other PLSA-based methods, which illustrates the validity of the sketch structural feature. Moreover, the black rectangle in the bottom right corner highlights the classification of buildings, and C-SSTM performs slightly better than SSTM since the structural constraint discovers the intricate correlation hidden in high-dimensional features. The deep learning methods perform well in road classification within the ellipse, because the neural network has an impressive feature representation ability for simple structure areas (such as road) through the multilayer feature extraction. However, within the rectangle, the deep learning method has unfavorable classification results in the region of the buildings. The network equally processes all pixels of image patches, and the number of pixels representing the structure is small. Disturbed by a large number of texture pixels, deep learning methods show an ill-structural representation ability. From the above analysis, it can be illustrated that the proposed method can yield stable and superb classification performance through sketch structural features and latent semantic topics.

As shown in the black ellipse of Fig. 14, a large number of residential areas are misclassified as commercial areas in the classification results of GTM, STM, and SALT M, while SSTM achieves a favorable classification results in the ellipse box. The residential classification results of C-SSTM appear to be more homogeneous than those via the other PLSA-based methods. In residential areas, the deep learning approaches perform well. The white ellipse of Fig. 14 highlights the region where the commercial areas are misclassified as the residential ones. The

classification accuracy of SSTM in white ellipse is comparable to those of GTM, STM, and SALT M. It can be observed that some misclassifications occur among the commercial and residential areas. There are two reasons for these misclassifications. First, these areas are composed of similar artificial buildings and have similarities in both the structure and texture features. Second, the size of sampled image patches is fixed single, the image patch cannot contain the complete structure when the object is large, and large amounts of irrelevant information may be contained in the patches when the object is small. In the white ellipse, the DCAE achieves unfavorable classification results since the image label information is not utilized in the training program. The classification accuracies obtained by CWNN and C-SSTM are equivalent in commercial areas. However, the performances of CWNN heavily rely on the number of network parameters learning to obtain satisfactory classification performance. In summary, the proposed method can achieve optimal performance with fewer computation costs.

In the black ellipse of Fig. 15, the classification results of SSTM and C-SSTM are better than those of the other three PLSA-based methods since the sketch structural features have an excellent capacity of feature separability. C-SSTM is shown to be good at discovering the intricate relationship between features in the manifold domain, further improving the classification accuracy of residential areas. It is observed that the deep learning methods achieve better classification accuracies than C-SSTM in the black ellipse. This is due to the fact that deep networks

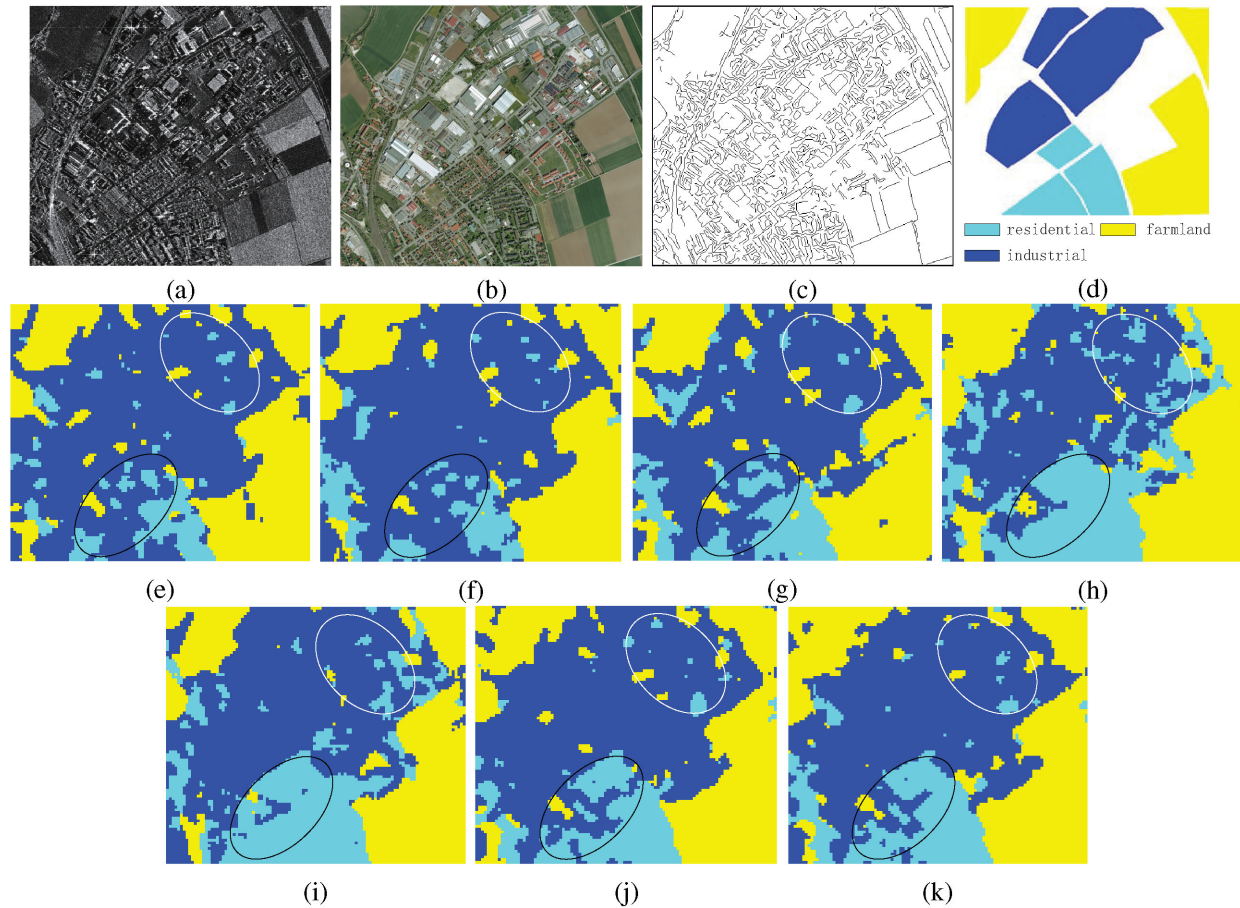


Fig. 15. Classification results of noerdlingernew. (a) Noerdlingernew. (b) Optical image. (c) Sketch map. (d) Ground truth. (e) GTM. (f) STM. (g) SALT M. (h) DCAE. (i) CWNN. (j) SSTM. (k) C-SSTM.

distinguish the similar spectrum categories with more network parameters. Meanwhile, some misclassification exists because several categories in the image are undefined, i.e., the idle land and road classes, which can appear in most terrain scenes. In the white ellipse, SSTM and C-SSTM get optimal classification accuracies in the industrial region. The deep learning methods yield quite misclassification pixels in the industrial area due to the complex spatial arrangement, spectral heterogeneity of industrial areas, and the insufficient extraction of structure features. In summary, the C-SSTM method can effectively extract the spatial and structural information, especially in the complex terrain areas. Making full use of the intricate relationship between different features in the manifold space, the C-SSTM increases the classification accuracy and regional consistency with a low computational cost.

The numerical results of the three real SAR images are listed in Tables VII–IX. The classification results of C-SSTM in complex land-cover areas are better than the other PTM-based methods, although the classification accuracy in some classes is slightly lower, but the average accuracy and Kappa coefficients are favorable, which indicate that the proposed method achieves better performance in most cases. Compared with the deep learning methods, the C-SSTM achieves a satisfied classification accuracy with less time consumption.

V. CONCLUSION

This article proposes a C-SSTM model based on the sketch structural feature and structural constraint for SAR image classification. The SAR image classification results depend on the discriminative representation and selection of features. Based on the sketch line segment to extract the key structure information at a higher semantic level, we designed the sketch structural feature extraction algorithm to obtain the structure feature. In terms of feature selection, different from the previous method using the local feature to learn the latent semantics, we obtain discriminative latent semantic topic features based on the prior knowledge of image manifold, combined with the overall texture and structure features into the PLSA model. Experimental results show that the proposed C-SSTM method performs excellent in discovering the discriminative semantic features from SAR images, with high time efficiency. However, the proposed algorithm still has limitations. Using the single image patch size may result in the disability to contain complete contextual semantic information. How to choose different sampling image patches is a problem to be considered. At the same time, the clustering algorithm plays a key role in processing image features. Improving the performance of clustering algorithms is another important research direction.

REFERENCES

- [1] Z. Xu, H.-C. Li, Q. Shi, H. Wang, and Y. Shao, "Effect analysis and spectral weighting optimization of sidelobe reduction on SAR image understanding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3434–3444, Sep. 2019.
- [2] F. Sharifzadeh, G. Akbarizadeh, and Y. S. Kaviani, "Ship classification in SAR images using a new hybrid CNN-MLP classifier," *J. Indian Soc. Remote. vol. 47, no. 4, pp. 551–562, 2019.*
- [3] F. Samadi, G. Akbarizadeh, and H. Kaabi, "Change detection in SAR images using deep belief network: A new training approach based on morphological images," *IET Image Process.*, vol. 13, no. 12, pp. 2255–2264, 2019.
- [4] P. Iervolino, R. Guida, D. Riccio, and R. Rea, "A novel multispectral, panchromatic and SAR data fusion for land classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3966–3979, Oct. 2019.
- [5] D. Guan, D. Xiang, X. Tang, L. Wang, and G. Kuang, "Covariance of textural features: A new feature descriptor for SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3932–3942, Oct. 2019.
- [6] T. L. M. Barreto *et al.*, "Classification of detected changes from multi-temporal high-RES Xband SAR images: Intensity and texture descriptors from superpixels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5436–5448, Dec. 2016.
- [7] C. O. Dumitru, S. Cui, G. Schwarz, and M. Datcu, "Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1635–1650, Apr. 2015.
- [8] J. Geng, H. Wang, J. Fan, and X. Ma, "Deep supervised and contractive neural network for SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2442–2459, Apr. 2017.
- [9] L. Chang and D. Y. Tsao, "The code for facial identity in the primate brain," *Cell*, vol. 169, no. 6, pp. 1013–1028, 2017.
- [10] M. E. J. Cutler, B. Boyd, G. M. Foody, and A. Vetrivel, "Estimating tropical forest biomass with a combination of SAR image texture and landsat TM data: An assessment of predictions between regions," *ISPRS J. Photogrammetry Remote Sens.*, vol. 70, pp. 66–77, Jun. 2012.
- [11] R. Ressel, A. Frost, and S. Lehner, "A neural network-based classification for sea ice types on X-band SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3672–3680, Jul. 2015.
- [12] G. Akbarizadeh, "A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4358–4368, Nov. 2012.
- [13] B. Hou, X. Zhang, and N. Li, "MPM SAR image segmentation using feature extraction and context model," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1041–1045, Nov. 2012.
- [14] S. Jia *et al.*, "3D Gaussian-Gabor feature extraction and selection for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8813–8826, Nov. 2019.
- [15] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang, "SAR image segmentation based on convolutional-wavelet neural network and Markov random field," *Pattern Recognit.*, vol. 64, pp. 255–267, 2017.
- [16] B. B. Saevansson, J. R. Sveinsson, and J. A. Benediktsson, "Speckle reduction of SAR images using adaptive curvelet domain," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2003, pp. 4083–4085.
- [17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110.
- [18] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2015.
- [19] V. Risojevic and Z. Babic, "Fusion of global and local descriptors for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 836–840, Jul. 2013.
- [20] Q. Zhu, Y. Zhong, S. Wu, L. Zhang, and D. Li, "Scene classification based on the sparse homogeneous-heterogeneous topic feature model," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2689–2703, May 2018.
- [21] Z. Tirandaz, G. Akbarizadeh, and H. Kaabi, "PolSAR image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization," *Measurement*, vol. 153, 2020, Art. no. 107432.
- [22] Z. Ren, B. Hou, Z. Wen, and L. Jiao, "Patch-sorted deep feature learning for high resolution SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3113–3126, Sep. 2018.
- [23] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, Apr. 2015.
- [24] J. Geng, H. Wang, J. Fan, and X. Ma, "SAR image classification via deep recurrent encoding neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2255–2269, Apr. 2018.
- [25] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, 2017.
- [26] M. Zalpour, G. Akbarizadeh, and N. Alaei-Sheini, "A new approach for oil tank detection using deep learning features with control false alarm rate in high-resolution satellite imagery," *Int. J. Remote Sens.*, vol. 41, no. 6, pp. 2239–2262, 2020.
- [27] Z. Ren, B. Hou, Q. Wu, Z. Wen, and L. Jiao, "A distribution and structure match generative adversarial network for SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3864–3880, Jun. 2020.
- [28] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
- [29] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state-of-the-art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [30] H. Fu, G. Sun, J. Zabalza, A. Zhang, J. Ren, and X. Jia, "A novel spectral-spatial singular spectrum analysis technique for near real-time in situ feature extraction in hyperspectral imaging," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2214–2225, 2020.
- [31] D. Bratasanu, I. Nedelcu, and M. Datcu, "Bridging the semantic gap for satellite image annotation and automatic mapping applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 193–204, Mar. 2011.
- [32] S. Xu, T. Fang, D. Li, and S. Wang, "Object classification of aerial images with bag-of-visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 366–370, Apr. 2010.
- [33] B. Hou, B. Ren, G. Ju, H. Li, L. Jiao, and J. Zhao, "SAR image classification via hierarchical sparse representation and multisize patch features," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 33–37, Jan. 2016.
- [34] Z. Zhao, L. Jiao, F. Liu, J. Zhao, and P. Chen, "Semisupervised discriminant feature learning for SAR image category via sparse ensemble," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3532–3547, Jun. 2016.
- [35] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Nice, France, 2003, pp. 1470–1477.
- [36] L.-J. Zhao, P. Tang, and L.-Z. Huo, "Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4620–4631, Dec. 2014.
- [37] H. M. Wallach, "Topic modeling: Beyond bag-of-words," in *Proc. 23rd Int. Conf. Mach. Learn.*, ACM, 2006, pp. 977–984.
- [38] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, no. 1–2, pp. 177–196, 2001.
- [39] C. He, T. Zhuo, D. Ou, M. Liu, and M. Liao, "Nonlinear compressed sensing-based LDA topic model for polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 3, pp. 972–982, Mar. 2014.
- [40] K. Than and T. B. Ho, "Fully sparse topic models," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, Springer, 2012, pp. 490–505.
- [41] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 712–727, Apr. 2008.
- [42] W. Yang, D. Dai, B. Triggs, and G.-S. Xia, "SAR-based terrain classification using weakly supervised hierarchical Markov aspect models," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4232–4243, Sep. 2012.
- [43] S. Wu, X. Jing, J. Yang, and J. Yang, "Learning image manifold using neighboring similarity integration," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 1897–1901.
- [44] J. Wu, F. Liu, L. Jiao, X. Zhang, H. Hao, and S. Wang, "Local maximal homogeneous region search for SAR speckle reduction with sketch-based geometrical kernel function," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5751–5764, Sep. 2014.
- [45] C. Shi, F. Liu, L. Li, L. Jiao, Y. Duan, and S. Wang, "Learning interpolation via regional map for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3417–3431, Jun. 2015.

- [46] Y. Shen, P. Guturu, and B. P. Buckles, "Wireless capsule endoscopy video segmentation using an unsupervised learning approach based on probabilistic latent semantic analysis with scale invariant features," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 1, pp. 98–105, Jan. 2012.
- [47] A. Baraldi and F. Parmiggiani, "An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 293–304, Mar. 1995.
- [48] R. M. Haralick *et al.*, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, no. 6, pp. 610–621, Nov. 1973.
- [49] U. Kandaswamy, D. A. Adjero, and M.-C. Lee, "Efficient texture analysis of SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2075–2083, Sep. 2005.
- [50] W. Luo, H. Li, G. Liu, and L. Zeng, "Semantic annotation of satellite images using author–genre–topic model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1356–1368, Feb. 2014.
- [51] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE 37th Asilomar Conf. Signals Syst. Comput.*, 2003, pp. 1398–1402.
- [52] Y. Yuan, X. Yao, J. Han, L. Guo, and M. Q.-H. Meng, "Discriminative joint-feature topic model with dual constraints for WCE classification," *IEEE Trans. Cybern.*, vol. 48, no. 7, pp. 2074–2085, Jul. 2018.
- [53] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution SAR image classification via deep convolutional autoencoders," *IEEE Geosci. Remote. Sens. Lett.*, vol. 12, no. 11, pp. 1–5, Nov. 2015.
- [54] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2108–2123, Apr. 2016.



Yake Zhang received the B.S. degree in mathematics and applied mathematics from Xuchang University, Xuchang, China, in 2012, and the M.S. degree in applied mathematics from Xidian University, Xi'an, China, in 2015. She is currently working toward the Ph.D. degree with the School of Artificial Intelligence, Xidian University.

Her research interests include machine learning and image processing.



Fang Liu (Senior Member, IEEE) received the B.S. degree in computer science and technology from Xi'an Jiaotong University, Xi'an, China, in 1984, and the M.S. degree in computer science and technology from Xidian University, Xi'an, China, in 1995.

She is currently a Professor with the School of Computer Science, Xidian University. Her research interests include signal and image processing, synthetic aperture radar image processing, multiscale geometry analysis, learning theory and algorithms, optimization problems, and data mining.

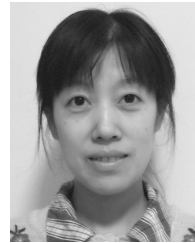


Licheng Jiao (Fellow, IEEE) received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees in electronic engineering from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

Since 1992, he has been a Professor with the School of Artificial Intelligence, Xidian University, Xi'an, China, where he is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China.

His research interests include image processing, natural computation, machine learning, and intelligent information processing.

Dr. Jiao is a member of the IEEE Xi'an Section Execution Committee. He is also the Chairman of the Awards and Recognition Committee, the Vice Board Chairperson of the Chinese Association of Artificial Intelligence, a Councilor of the Chinese Institute of Electronics, a Committee Member of the Chinese Committee of Neural Networks, and an Expert of the Academic Degrees Committee of the State Council.



Shuyuan Yang (Senior Member, IEEE) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in circuit and system from Xidian University, Xi'an, China, in 2000, 2003, 2005, and 2010, respectively.

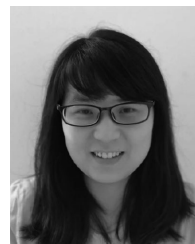
Since 2010, she has been a Professor with the School of Artificial Intelligence, Xidian University. Her research interests include machine learning and multiscale geometric analysis.



Lingling Li (Member, IEEE) received the B.S. degree in electronic information engineering and the M.S. and Ph.D. degrees in intelligent information processing from Xidian University, Xi'an, China, in 2011, 2012, and 2017, respectively.

From 2013 to 2014, she was an Exchange Ph.D. Student with the Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Leioa, Spain. She is currently a Postdoctoral Researcher with the School of Artificial Intelligence,

Xidian University. Her research interests include quantum evolutionary optimization, machine learning, and deep learning.



Meijuan Yang (Graduate Student Member, IEEE) received the B.S. degree from Xidian University, Xi'an, China, in 2010, and the M.S. degree in signal and information processing from the Xi'an Institute of Optics and Precision Mechanics, University of Chinese Academy of Sciences, Xi'an, China, in 2013. She is currently working toward the Ph.D. degree with the School of Artificial Intelligence, Xidian University.

Her research interests include deep learning and image processing.