

Article

Scene Classification from Synthetic Aperture Radar Images Using Generalized Compact Channel-Boosted High-Order Orderless Pooling Network

Kang Ni, Yiquan Wu * and Peng Wang

School of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Jiang Jun Road No. 29 Jiangjun Ave., Nanjing 211100, China; baiyaruibigbang@nuaa.edu.cn(K.N.); pengwang_b614080003@nuaa.edu.cn(P.W.)

* Correspondence: imagestrong@nuaa.edu.cn; Tel.: +025-84896490-4304

Received: 1 April 2019; Accepted: 4 May 2019; Published: 7 May 2019



Abstract: The convolutional neural network (CNN) has achieved great success in the field of scene classification. Nevertheless, strong spatial information in CNN and irregular repetitive patterns in synthetic aperture radar (SAR) images make the feature descriptors less discriminative for scene classification. Aiming at providing more discriminative feature representations for SAR scene classification, a generalized compact channel-boosted high-order orderless pooling network (GCCH) is proposed. The GCCH network includes four parts, namely the standard convolution layer, second-order generalized layer, squeeze and excitation block, and the compact high-order generalized orderless pooling layer. Here, all of the layers are trained by back-propagation, and the parameters enable end-to-end optimization. First of all, the second-order orderless feature representation is acquired by the parameterized locality constrained affine subspace coding (LASC) in the second-order generalized layer, which cascades the first and second-order orderless feature descriptors of the output of the standard convolution layer. Subsequently, the squeeze and excitation block is employed to learn the channel information of parameterized LASC statistic representation by explicitly modelling interdependencies between channels. Lastly, the compact high-order orderless feature descriptors can be learned by the kernelled outer product automatically, which enables low-dimensional but highly discriminative feature descriptors. For validation and comparison, we conducted extensive experiments into the SAR scene classification dataset from TerraSAR-X images. Experimental results illustrate that the GCCH network achieves more competitive performance than the state-of-art network in the SAR image scene classification task.

Keywords: synthetic aperture radar; scene classification; convolutional neural network; orderless pooling; feature representations

1. Introduction

With the rapid development of various synthetic aperture radar (SAR) sensors, large amounts of high-quality SAR remote sensing images have been produced. SAR images have the advantages of a certain penetration ability of vegetation and work in all weathers and round-the-clock [1]. These images have been widely used in land cover classification [2], marine observation [3,4], and environment projection [5]; however, characterizing SAR images effectively is still one of the most challenging tasks in these applications.

SAR image scene classification, which is a direct understanding and interpretation of SAR data [6,7], has become an important technique in the extraction of the ground objects, target recognition and so on [8]. The description of SAR images is more difficult than high-resolution optical remote sensing

images. As shown in the Figure 1, speckle noise makes the extraction of discriminative descriptors difficult in SAR scene images; in particular, there are many irregular repetitive patterns which are spatially correlated in the SAR image [6]. Additionally, various kinds of objects exhibit diversified singularities which have different shapes, contours and textures in the scene image. Considering all of them, SAR image scene classification is indeed onerous.

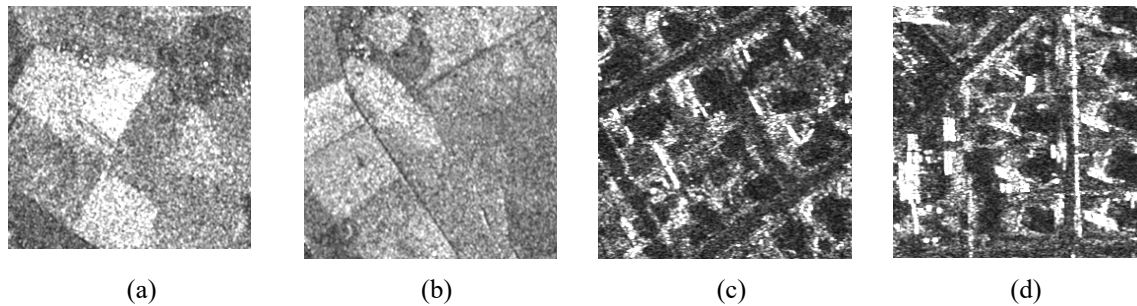


Figure 1. The synthetic aperture radar (SAR) scene images in the same class. (a) and (b) are the agricultural images, (c) and (d) are the building images.

Inspired by the great success achieved by CNN [9] in the computer vision community, the powerful feature representations learnt through CNN and the parameters enabling end-to-end optimization [10] make the characterization of SAR scene images possible. Research shows the different layers learn different levels of semantic information in the CNN [11,12]; they extract high-level semantic information from the edges and corners which are most salient in the deepest layer [13], and this high-level semantic information is helpful for SAR image scene classification. However, the standard convolution operation is performed by a fixed-size convolution kernel with a fixed step size corresponding to the feature map of the previous layer, and the obtained calculation result is also placed at a fixed position of the output; this is the same as the standard pooling operation. This has led to deep feature representation having strong global spatial information, and the geometric invariance is weak. To improve the performance of SAR remote sensing scene classification, we should globally order the spatial structure in small regions, and the spatial structure should be orderless in large regions (due to the layout differences of the scene image); however, the deep feature representation lacks the orderless feature descriptor [14]. The bag of visual words (BoVW) model [15–17] which produces orderless features and makes features characterize intra-class differences [18] achieves better performance on the scene classification. The famous feature coding used in the BoVW pipeline includes vector quantization (VQ) [19,20], sparse coding (SC) [21], locality-constrained linear coding (LLC) [22], super vector (SV) [23], Fisher vector (FV) [24], locality-constrained affine subspace coding (LASC) [25], and so on. All of them could generate orderless feature representation, which is helpful for improving the accuracy of scene classification.

Nowadays, more and more experts are focused on the BoVW model [26] and CNN [27] for remote sensing scene classification [28–30]. Cheng. et al [31] present a novel feature representation method for scene classification based on the bag of convolutional features (BoCF) which generates visual words from deep convolutional features using convolutional neural networks. Yuan. et al [32] obtain more discriminative deep features than direct extraction from the top convolutional layers of CNN, and the final feature representations are obtained by the LASC pooling of deep features. Meanwhile, Banerjee. et al [33] present a pattern mining-based approach for the efficient discovery of mid-level visual elements, which highlights the correlations between such local descriptors and the classes efficiently. Although these works have shown their powerful feature representation in remote sensing scene classification, they still have some shortcomings: (1) only the first-order orderless feature is considered, while high-order orderless feature representation, which can better characterize the distributions of SAR scene images, is missed; (2) the parameters of the BoVW model do not enable end-to-end optimization to make the model less flexible; and (3) the channel information of the

feature descriptor is fixed in the BoVW model to make the feature descriptors less discriminative for the classification.

To solve the above issues, a generalized compact channel-boosted high-order orderless pooling network (GCCH) is presented. The proposed GCCH network can generate a compact channel-boosted high-order orderless feature descriptor, which is low-dimensional but has high discrimination; additionally, the parameters enable end-to-end optimization. The overall framework of the proposed network is illustrated in Figure 2.

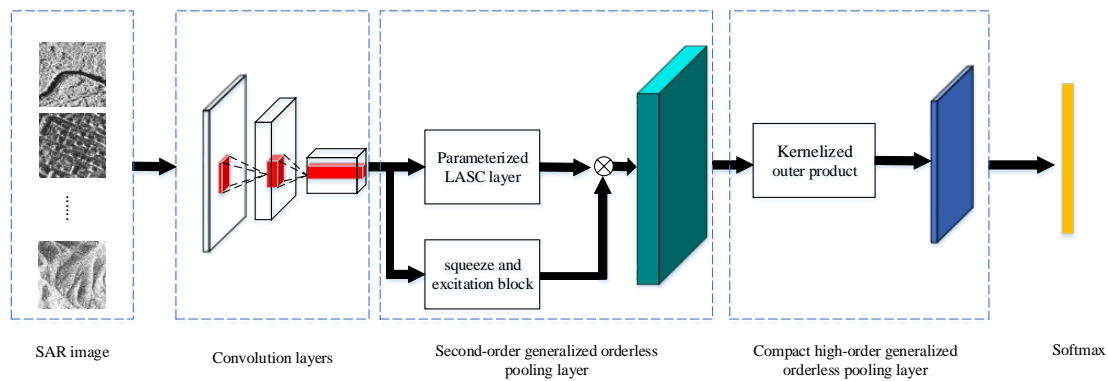


Figure 2. The overall framework of the generalized compact channel-boosted high-order orderless pooling (GCCH) network. The proposed GCCH network includes the standard convolution layers, second-order generalized layer (parameterized locality constrained affine subspace coding (LASC) layer), squeeze and excitation block, and the compact high-order generalized orderless pooling layer. The black arrow represents the forward operation of the proposed network; all of the parameters can be updated by back-propagation in the GCCH network.

The main contributions of this paper are summarized below.

- We present a generalized compact channel-boosted high-order orderless pooling network (GCCH) which learns the high-order vector of parameterized locally constrain affine subspace coding by a kernelled outer product. It enables low-dimensional but highly discriminative feature descriptors, and the dimension of the final feature descriptors can be set by us.
- We employ the squeeze and excitation block to learn the channel information of the parameterized LASC statistic representation by explicitly modelling the interdependencies between channels; the block obtains the importance of each feature channel by the adaptive learning network.
- All of the layers can be trained by back-propagation, and some important parameters on the proposed network are made to obtain better experimental parameter settings, and research on the SAR image scene classification dataset demonstrates the competitive performance.

The remainder of the paper is organized as follows: Section 2 discusses the traditional LASC method; the generalized compact channel-boosted high-order orderless pooling network is proposed in Section 3; comprehensive experimental results are reported in Section 4; and Section 5 concludes the paper and proposes future research work.

2. Traditional LASC Method

In the past decade, the bag of visual words model has played an important role in a variety of vision tasks. The standard BoVW pipeline [34,35] consists of feature extraction, dictionary learning, feature coding and spatial aggregation, and classification. Locality-constrained affine subspace coding (LASC), which is part of the BoVW model, performs first-order coding over a dictionary of affine subspaces and second-order coding based on information geometry, which has proven helpful to improve classification accuracy. Figure 3 gives the illustration of the traditional LASC method.

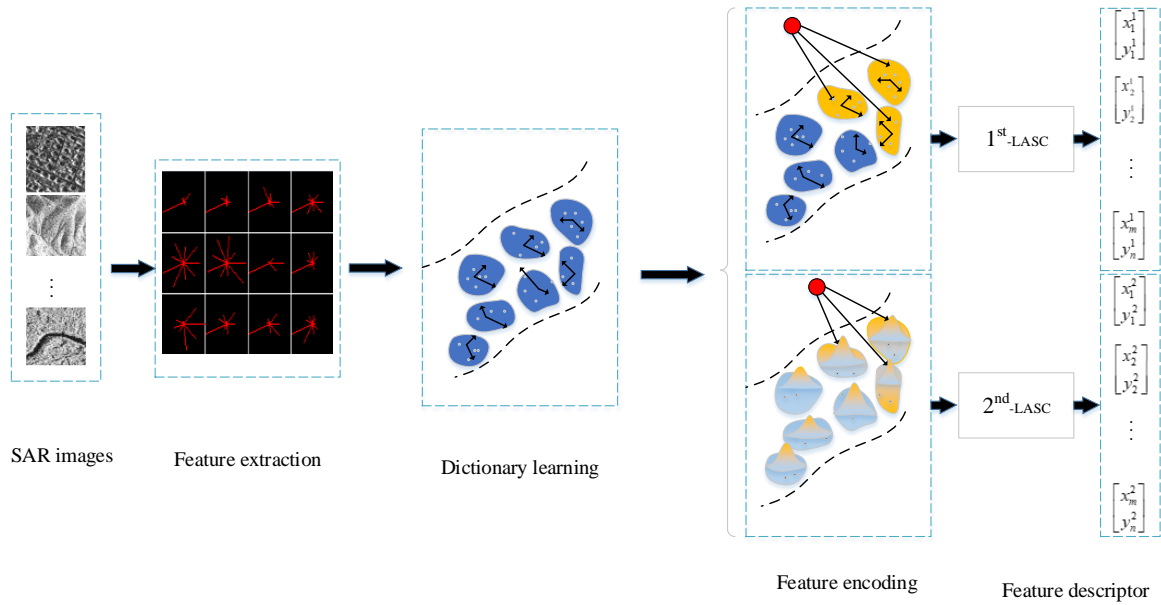


Figure 3. The illustration of the traditional LASC method. The LASC features are extracted by hand crafting, such as the histogram of oriented gradient (HOG) [36], and the dictionary of LASC is an ensemble of low-dimensional linear subspaces attached to the affine subspaces.

In the traditional LASC method, the geometry of the feature space is represented by an ensemble \mathbf{S} of low-dimensional subspace \mathbf{S}_i [25],

$$\mathbf{S}_i = \{\mu_i + \mathbf{A}_i c_i, c_i \in R^p\}, i = 1, \dots, M \tag{1}$$

where μ_i is the attachment point of the i th subspace, which is the cluster center obtained by the K -means algorithm. The columns of \mathbf{A}_i ($D \times p$ matrix, D is the feature dimension, p is the dimension of the subspace) form an orthogonal basis of the linear subspace. \mathbf{S}_i defines a local coordinate system, and all of \mathbf{S}_i constitutes an approximation of the feature distribution actually.

The principle of LASC encoding for feature y is to express the nearest top- k affine subspace, while using the neighboring metric constraint feature to project the projection vector in each neighbor affine subspace. Then, the objective function of LASC is formulated as

$$\min_{\forall c_i} \sum_{S_i \in N_k^S(y)} \left\{ \|(y - \mu_i) - c_i\|_2^2 + \lambda d(y, S_i) \|c_i\|_2 \right\} \tag{2}$$

where $\lambda > 0$ is the parameter of the regular item. Let $N_k^S(y)$ be the set of k , which is the closest subspaces of y ; $d(y, S_i)$ denotes the proximity measure between y and S_i , and $\|\bullet\|_2$ denotes the ℓ_2 -norm. Based on reconstruction error, $d(y, S_i)$ is defined as

$$d(y, S_i) = \frac{\exp(-\gamma \|(y - \mu_i) - \mathbf{A}_i c_i\|_2^2)}{\sum_{i' \in N_k^S(y)} \exp(-\gamma \|(y - \mu_{i'}) - \mathbf{A}_{i'} c_{i'}\|_2^2)} \tag{3}$$

where $d(y, S_i)$ is defined as the exponentiated Euclidean distance, and γ is obtained by cross-validation. We solve Equation (2) and obtain the closed-form solution of LASC, named c_i . Finally, cascading the feature coding of each subspace, its first-order LASC vector takes the following form:

$$\mathbf{c} = [\omega(y, \mathbf{S}_1) c_1^T, \dots, \omega(y, \mathbf{S}_i) c_i^T, \dots, \omega(y, \mathbf{S}_M) c_M^T] \tag{4}$$

where . In particular, we can derive the second-order coding form of LASC by the Fisher information, then define the gradient vector of the likelihood function as

$$g_{S_i} = \nabla_{S_i} \log p(z_i|S_i) \tag{5}$$

where $p(z_i|S_i)$ is the log-likelihood function of z_i , and then the Fisher information metric is

$$G_{S_i} = E_{p(z|S_i)}[g_{S_i}g_{S_i}^T] \tag{6}$$

where $E(\bullet)$ denotes expectation with respect to $p(z_i|S_i)$. Lastly, based on Fisher’s information theory, the Fisher vector of z_i is defined as

$$f_{S_i} = G_{S_i}^{-1/2}g_{S_i} = \frac{1}{\sqrt{2}} \left[\frac{(z_{i,1})^2}{\beta_{i,1}} - 1, \dots, \frac{(z_{i,1})^2}{\beta_{i,\rho}} - 1 \right] \tag{7}$$

The second-order LASC is calculated as a weighted Fisher vector, which is formulated as

$$f = \left[\omega(\mathbf{y}, \mathbf{S}_1)f_{S_1}^T, \dots, \omega(\mathbf{y}, \mathbf{S}_i)f_{S_i}^T, \dots, \omega(\mathbf{y}, \mathbf{S}_M)f_{S_M}^T \right]^T \tag{8}$$

if $S_i \notin N_k^s(\mathbf{y})$, $f_{S_i} = 0$. Finally, we concatenate the first and second-order LASC statistical information, and we will acquire the entire LASC feature representation.

3. Generalized Compact Channel-Boosted High-Order Orderless Pooling Network

In this section, the proposed GCCH network will be presented in detail. As illustrated in Figure 2, the proposed network includes four parts: (1) the standard convolution layers, (2) the second-order generalized layer (parameterized LASC layer), (3) the squeeze and excitation block, and (4) the compact high-order generalized orderless pooling layer. Table 1 gives the differences between the GCCH network and other approaches. The GCCH network learns the powerful high-order deep feature vector and the channel information by end-to-end training; additionally, the low-dimensional feature descriptor is produced to reduce the redundancy of the high-order feature vector.

Table 1. The differences between the proposed GCCH network and other approaches.

Method	Include Layers	Channel Learning	High-Order Feature	Parameters Learning	Feature	End to end Training
Traditional LASC [25]	\	\	No	No	hand crafted	No
BoCF [31]	\	\	No	No	deep feature	No
NetVLAD [37]	VLAD [38] (first-order)	No	No	Yes	deep feature	Yes
GCCH (Ours)	LASC (second-order)	Yes	Yes	Yes	deep feature	Yes

3.1. The Second-Order Generalized Layer (Parameterized Lasc Layer)

First of all, the closed-form solution of Equation (2) takes the following form:

$$c_i = \left(A_i^T A_i + \lambda I \right)^{-1} A_i^T (y - \mu_i) \tag{9}$$

where $\left(A_i^T A_i + \lambda I \right)^{-1}$ is the hard assignment coefficient; according to Equation (3), we replace it with a soft assignment coefficient [37,39],

$$\kappa_i = \frac{e^{-\beta \|y_i - \mu_i\|_2^2}}{\sum_{k'=1}^K e^{-\beta \|y_i - \mu_{k'}\|_2^2}} \tag{10}$$

where β is a positive constant, K is the number of visual words (cluster centers [40]). After expanding the squares, the soft assignment coefficient can be expressed as

$$\kappa_i = \frac{e^{w_k^T y_i + b_k}}{\sum_{k'=1}^K e^{w_{k'}^T y_i + b_{k'}}} \quad (11)$$

where $w_k = 2\beta\mu_i$, and $b_k = -\beta\|\mu_i\|^2$. LASC is a localized and second-order coding method and only considers the nearest top-k affine subspace. The soft-assignment coefficient of LASC layer is given by

$$\kappa_i = \begin{cases} \frac{e^{w_k^T y_i + b_k}}{\sum_{k' \in N_T(y_i)} e^{w_{k'}^T y_i + b_{k'}}}, & k \in N_T(y_i) \\ 0 & , \text{ otherwise} \end{cases} \quad (12)$$

where $N_T(y_i)$ (T nearest neighbor layer, TNN layer) denotes the indexes of T nearest neighbor visual word of y_i which is based on Equation (3). Meanwhile, the dimension of dictionary A is reduced by the affine subspace layer which is based on the nearest neighbor principal component analysis (PCA) basis, and the projective matrix is defined by P_i ($P_i \in R^{S \times D}$), and S is the dimensionality of the subspace. The forward computation of the first-order LASC feature representation is written as

$$L^1 = \sum_{i=1}^M \kappa_i P_i (y_i - \mu_i) \quad (13)$$

where M denotes the number of the feature descriptors. Different from NetVLAD and LLC, LASC utilizes the second-order statistical information, according to Equation (7), and after a series of simple transformations, the forward computation of the second-order LASC statistical information is given by

$$L^2 = \sum_{i=1}^M \kappa_i [(P_i (y_i - \mu_i))^2 - 1] \quad (14)$$

We will acquire the final LASC feature representation by cascading both L^1 and L^2 ; the network architecture of parameterized LASC layer is shown in Figure 4.

Firstly, the deep feature is normalized by l_2 -Norm (element-wise normalization). According to Equation (11), $w_k^T y_i + b_k$ can be viewed as the 1×1 convolution layer with the weight $\{w\}$ and bias $\{b\}$ which are well known in CNN; the exponential layer is employed to calculate $e^{w_k^T y_i + b_k}$, which is similar to the softmax layer. The TNN layer only considers the T nearest neighbor visual word of y_i ; if it is other than 0, we can implement it by the variant of the max pooling layer. The sum-norm layer is calculated by Equation (12).

The affine subspace layer is defined as the 1×1 convolution layer with the weight $\{P\}$ and bias $\{-P \times \mu\}$, which is based on $P_i (y_i - \mu_i)$ in Equation (13), and we will acquire the first-order LASC feature representation through the product of the soft assignment layer and affine subspace layer. As seen in Equations (13) and Equation (14), the second-order LASC can be viewed as the nonlinear active function of first-order LASC; we define the activation function as $\varphi(x) = x^2 - 1$ in the activation layer, and the derivative of $\varphi(x)$ can be easily obtained. Then, the second-order statistic information is obtained. Lastly, we cascade them to acquire the final LASC statistic information. Here, all of the parameters ($\{w\}$, $\{b\}$, $\{P\}$ and $\{-P \times \mu\}$) can be trained by back-propagation in the parameterized LASC layer. Then, the algorithm steps (forward computation) of the parameterized LASC layer are given as follows:

Input: the output of the convolution u

Output: the entire LASC statistic information

- Step1. Normalize u by the l_2 -norm layer.
 Step2. The first-order LASC static information L^1 is obtained by rescaling the output of the soft assignment layer with the affine subspace layer.
 Step3. The second-order LASC static information L^2 is obtained by rescaling the output of the soft assignment layer with the affine subspace layer and activation layer.
 Step4. Cascading both L^1 and L^2 .

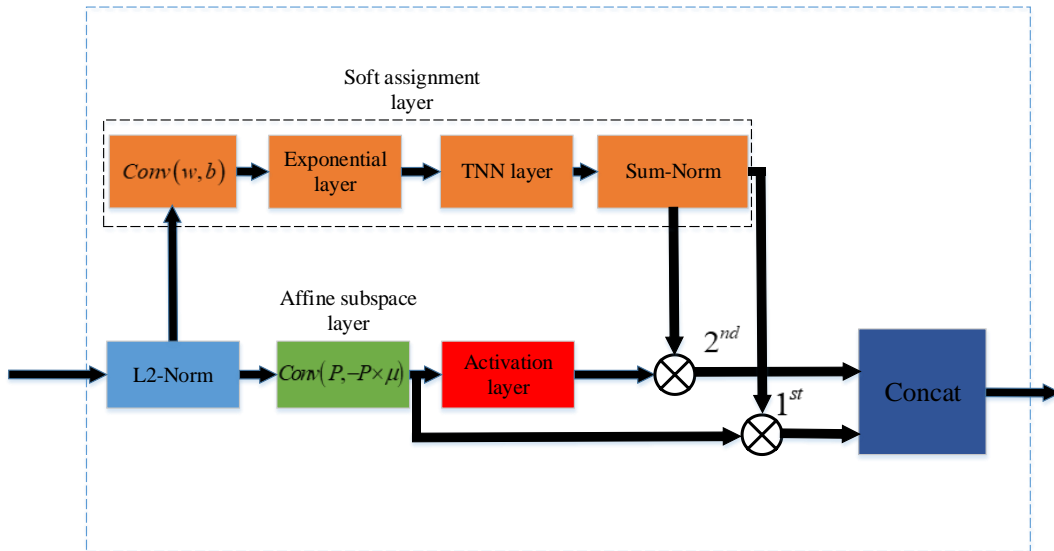


Figure 4. The network architecture of the parameterized LASC layer. The input of the parameterized LASC layer is the standard convolution layer in the convolutional neural network (CNN), which is a D dimensional feature vector. Both the dimensionality of the first-order and second-order LASC static information are $K \times S$. Finally, we cascade them.

3.2. Squeeze and Excitation Block

The channel information of the feature descriptor is fixed in the popular BoVW model; this makes the feature descriptors less discriminative for the classification. In order to overcome this shortcoming, the squeeze and excitation block is employed to learn the channel information of each second-order feature statistic representation [41] by explicitly modelling the interdependencies between channels; it can obtain the importance of each feature channel by an adaptive learning network. According to this, we should enhance the powerful features and suppress the features which are not helpful for SAR image scene classification. Figure 5 reports the schematic of the squeeze and excitation block.

The squeeze operation is used to generate the channel-wise statistics by using the global average pooling operation after the convolution layer, so that the feature maps ($H \times W \times C$) become the real number series of $1 \times 1 \times C$. z is defined as the output of squeeze operation, and the c^{th} element of z is given by

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (15)$$

where u denotes the output of the convolution layer. The convolution layer can be viewed as the collection of local descriptors, and the global average pooling layer makes the descriptors have a global receptive field.

The excitation operation employs the simple gating mechanism to learn the channel-wise statistics of features automatically. The excitation operation employs two fully connected (FC) layers [42]

(both of which are 1×1 convolution layers) and an ReLU activation layer to limit the complexity and generalization of the model, which can be shown as

$$e = \sigma(W_2\delta(W_1z)) \tag{16}$$

The first FC layer is named the dimensionality-reduction layer [43] ($W_1 \in \mathbb{R}^{(C/r) \times C}$, r is ratio of dimension reduction), δ is the ReLU activation function, the second FC layer is named the dimensionality increasing layer with parameters W_2 ($W_2 \in \mathbb{R}^{C \times (C/r)}$) in the Figure 5, and σ is the sigmoid activation function and obtains the final channel-wise statistics ($1 \times 1 \times C$). The algorithm steps (forward computation) of the squeeze and excitation block are illustrated as follows:

Input: the output of the convolution u

Output: the channel information of parameterized LASC statistic representation

- Step1. The squeeze operation z is calculated by Equation (15).
- Step2. The excitation operation e is obtained by Equation (16).
- Step3. Then, the output ($H \times W \times C$) is obtained by rescaling the convolution output u with e .

The squeeze and excitation block learns the channel information of the parameterized LASC statistic representation by explicitly modelling the interdependencies between channels, and helps to boost feature discriminability.

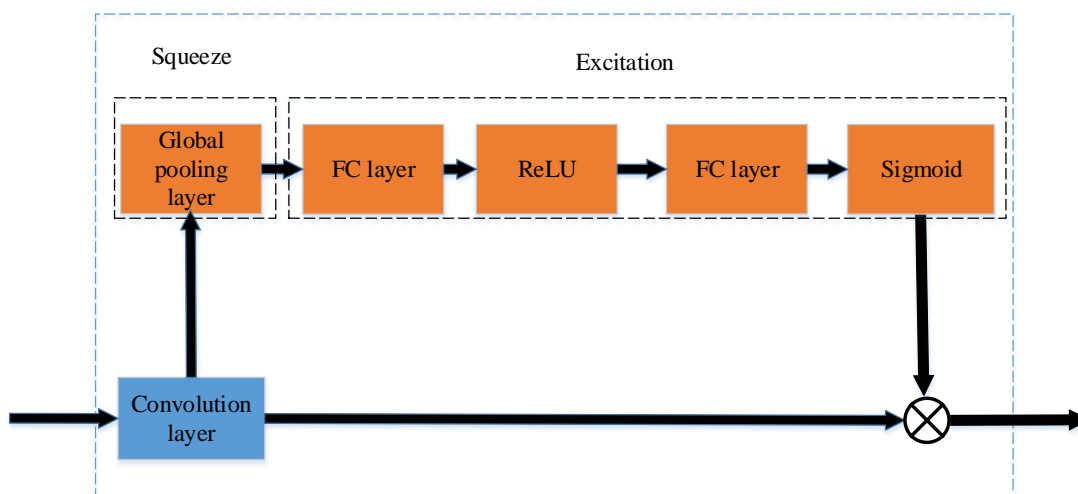


Figure 5. The schematic of squeeze and excitation block, including the squeeze operation and excitation operation.

3.3. Compact High-Order Generalized Orderless Pooling Layer

The high-order orderless feature descriptor is obtained by the kernelled outer product at a single spatial location; compared with the outer product [44,45], it enables low-dimensional but highly discriminative feature descriptors. Given the input of the compact high-order generalized orderless pooling layer $X = (x_1, \dots, x_{|S|}, x_s \in \mathbb{R}^c)$, S is the set of spatial locations, and the fully bilinear pooling ($c \times c$ matrix) can be shown as

$$\begin{aligned} \langle B(X), B(X) \rangle &= \left\langle \sum_{s \in S} x_s x_s^T, \sum_{s \in S} x_s x_s^T \right\rangle \\ &= \sum_{s \in S} \langle x_s x_s^T, x_s x_s^T \rangle \\ &= \sum_{s \in S} \langle x_s, x_s \rangle^2 \end{aligned} \tag{17}$$

where $\langle \bullet \rangle$ is the inner product operation; the fully bilinear descriptors are a second-order polynomial kernel, and it is possible to find low-dimensional bilinear descriptors. Let $k(x, x)$ denote the kernel function, and $\langle \phi(x), \phi(x) \rangle \approx k(x, x)$, $\phi(x) \in \mathbb{R}^d$, $d < c^2$. Equation (17) can be approximated by

$$\begin{aligned} \langle B(X), B(X) \rangle &= \sum_{s \in S} \langle x_s, x_s \rangle^2 \\ &\approx \sum_{s \in S} \langle \phi(x), \phi(x) \rangle \\ &= \langle C(X), C(X) \rangle \end{aligned} \quad (18)$$

where $C(X)$ is the compact high-order orderless feature descriptor; we utilize the random maclaurin (RM) projection approach [46] to approximate these operations. The forward computation of the compact high-order generalized orderless pooling layer is given in the following steps:

Input: the output of second-order generalized orderless pooling layer x

Output: the compact high-order orderless feature representation $\phi(x)$

Step1Generate random $W^1, W^2 \in \mathbb{R}^{d \times c}$ (the parameters needed to be learned), and make each entry be either +1 or -1 with equal probability.

Step2Let $\phi(x) = \frac{1}{\sqrt{d}}(W^1 x) \circ (W^2 x)$, and \circ denotes element-wise multiplication.

Especially, the back-propagation of the compact high-order generalized orderless pooling layer takes the following form:

$$\begin{aligned} \frac{\partial L}{\partial x_s^n} &= \sum_d \frac{\partial L}{\partial y_d^n} \sum_k \langle W_k(d), x_s^n \rangle W_{k'}(d) \\ \frac{\partial L}{\partial W_k(d)} &= \sum_n \frac{\partial L}{\partial y_d^n} \sum_s \langle W_{k'}(d), x_s^n \rangle x_s^n \end{aligned} \quad (19)$$

where L denotes the loss function, d means the projected dimension, n is the index of the training sample, S is the spatial index, y_d^n denotes the output of the compact high-order generalized orderless pooling layer at dimension d —for instance n . $k = 1, 2$, $k' = 2, 1$ —and $W_k(d)$ is the d^{th} row of W_k . Here, the compact high-order generalized orderless pooling layer can be trained by the chain rule, and the parameters can be updated by back-propagation.

4. Experimental Results and Discussion

The challenging SAR dataset used for evaluating the proposed network is first given in this section. Subsequently, the experimental setup and experimental results for SAR image scene classification are discussed.

4.1. Data Set

The evaluation of the GCCH network is tested on the TerraSAR-X database; this dataset contains 5000 scene images from 10 classes (agricultural, beach, building, forest, intersection, island, mountain, ocean, river, and runway), and each class consists of 500 images with the size of 256×256 . Example images from the 10-class TerraSAR-X dataset are shown in Figure 6.

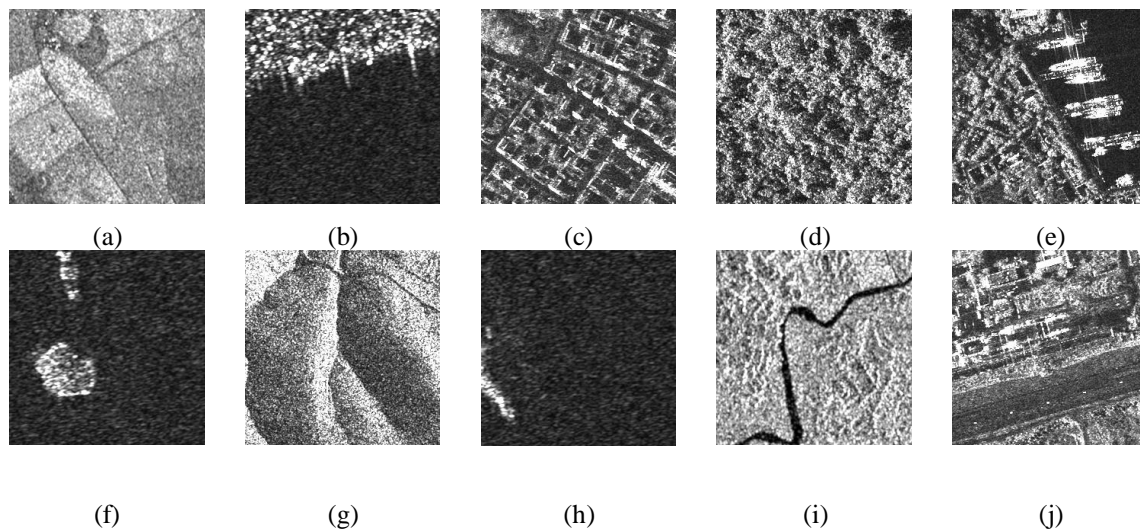


Figure 6. Example images from the TerraSAR-X dataset: (a) agricultural; (b) beach; (c) building; (d) forest; (e) intersection; (f) island; (g) mountain; (h) ocean; (i) river; (j) runway.

4.2. Experimental Setup

To demonstrate the effectiveness of the GCCH network, we truncate the conv5_3 layer of VGG-16 net as the standard convolution layers, which can be downloaded from <http://www.vlfeat.org/matconvnet/pretrained>. The parameters of the LASC layer are set as follows: the dimension of each subspace (subspace dimension) is set to be 256, the number of subspaces (dictionary size) is fixed at 128, and $k = 5$ (the number of nearest subspaces). The simulations of all the methods were carried out in Matlab R2016b on the PC with Intel Core i7-8700K @3.70GHz/16.00 GB RAM and NVIDIA 1070ti GPU; the average results are given in the following section.

For the TerraSAR-X dataset, we set the ratios of the training set to 10%, 20% and 30% (which is randomly selected per class), and 90%, 80% and 70% for testing.

4.3. Experimental Results and Discussion

First of all, a series of experimental results are made to analyze the influence of some important parameters on the GCCH network, and then we make comparisons with other state-of-the-art and mid-level feature representation approaches on the TerraSAR-X dataset.

4.3.1. Evaluation of Projected Dimensions

The GCCH network generates the high-order orderless feature representations which are parameterized by the user-defined projected dimensions d . We perform experiments on the TerraSAR-X dataset to investigate the parameter d of the proposed network. The implementation details are as follows: the ratio of reduction $r = 8$, the batch size is set to 16, the learning rates are fixed at 0.001, the momentum is set to 0.9, and the weight decay is 0.005. The Adam optimizer was employed throughout our experiments, and fine-tuned the parameters of layers from Conv4_3 and above in the proposed network. The experimental results are illustrated in Figure 7.

As shown in Figure 7, when $d = 128$ or 256 or 512, the experimental results decrease significantly using the low-dimensional representation, when $d \geq 1024$, the proposed network achieves better performance than the low-dimensional representation. Additionally, $d = 8192$ and $d = 16384$ get the highest score. In conclusion, considering the tradeoff between the model complexity and classification accuracy, we suggest that between 1000 and 8000 feature dimensions is appropriate, and we set $d = 8192$ in our paper.

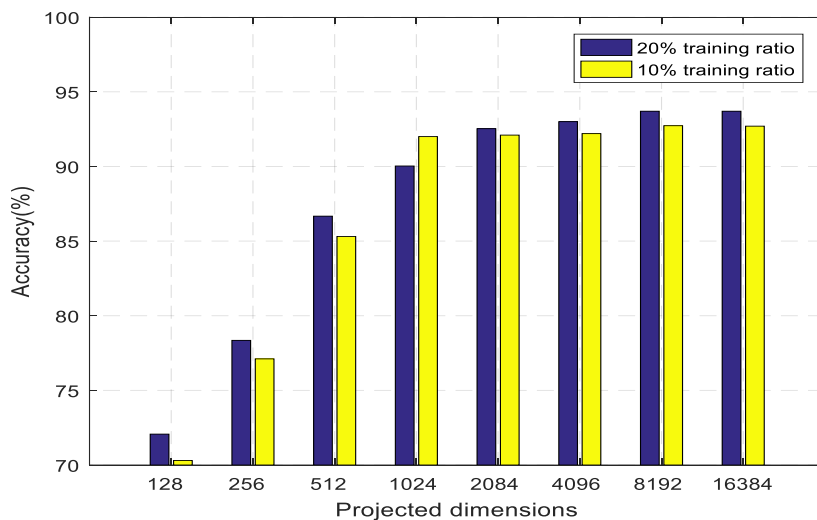


Figure 7. Influence of projected dimensions in the GCCH network.

4.3.2. Evaluation of the Ratio of Reduction

The ratio of reduction r affects the feature channel learning of the parameterized LASC layer and the parameters of the final model; different r brings different model complexity, and we set the ratio of reduction from 2 to 16 in the proposed network and make experiments on the TerraSAR-X dataset. The other parameters are set as follows: projected dimensions $d = 8192$, while the batch size, learning rates, momentum and weight decay are set the same as Section 4.3.1, and we fine-tune the parameters of layers from Conv4_3 and above in the GCCH network.

In the squeeze and excitation block, the additional parameters result solely from the two FC layers and make the total network capacity increase; however, the parameters will also affect the final model accuracy. Checking in Table 2, $r = 8$, $r = 4$ and $r = 2$ achieve similar performances which are better than $r = 16$, and the accuracy of $r = 16$ has a small degree of reduction. Approximately, the parameter of $r = 4$ is twice that of $r = 2$, and the parameter of $r = 8$ is about four times than that of $r = 2$. Here, considering the tradeoff between improved performance and increased model complexity, $r = 8$ is optimal in the GCCH network.

Table 2. Evaluation of the ratio of reduction.

r	TerraSAR-X Dataset		
	10% Training Ratio	20% Training Ratio	Parameters
2	92.73%	93.75%	395K
4	92.71%	93.73%	198K
8	92.70%	93.70%	100K
16	92.31%	92.89%	50K

4.3.3. Comparison with Other State-of-the Art Methods

In this subsection, we make comparisons with other state of the art methods on the TerraSAR-X dataset; all the competing methods are end-to-end training (based on VGG-16 net), such as B-CNN [44], NetVLAD [37], and LSO-VLADNet [39]. For the TerraSAR-X dataset, three different training-test ratios (10%-90%, 20%-80% and 30%-70%) are set; additionally, the projected dimensions $d = 8192$, the ratio of reduction $r = 8$, and the other parameters are set the same as in Section 4.3.1. Figure 8 presents the performance comparisons, and Table 3 gives the performance results (average accuracy) compared with other state of the art methods.

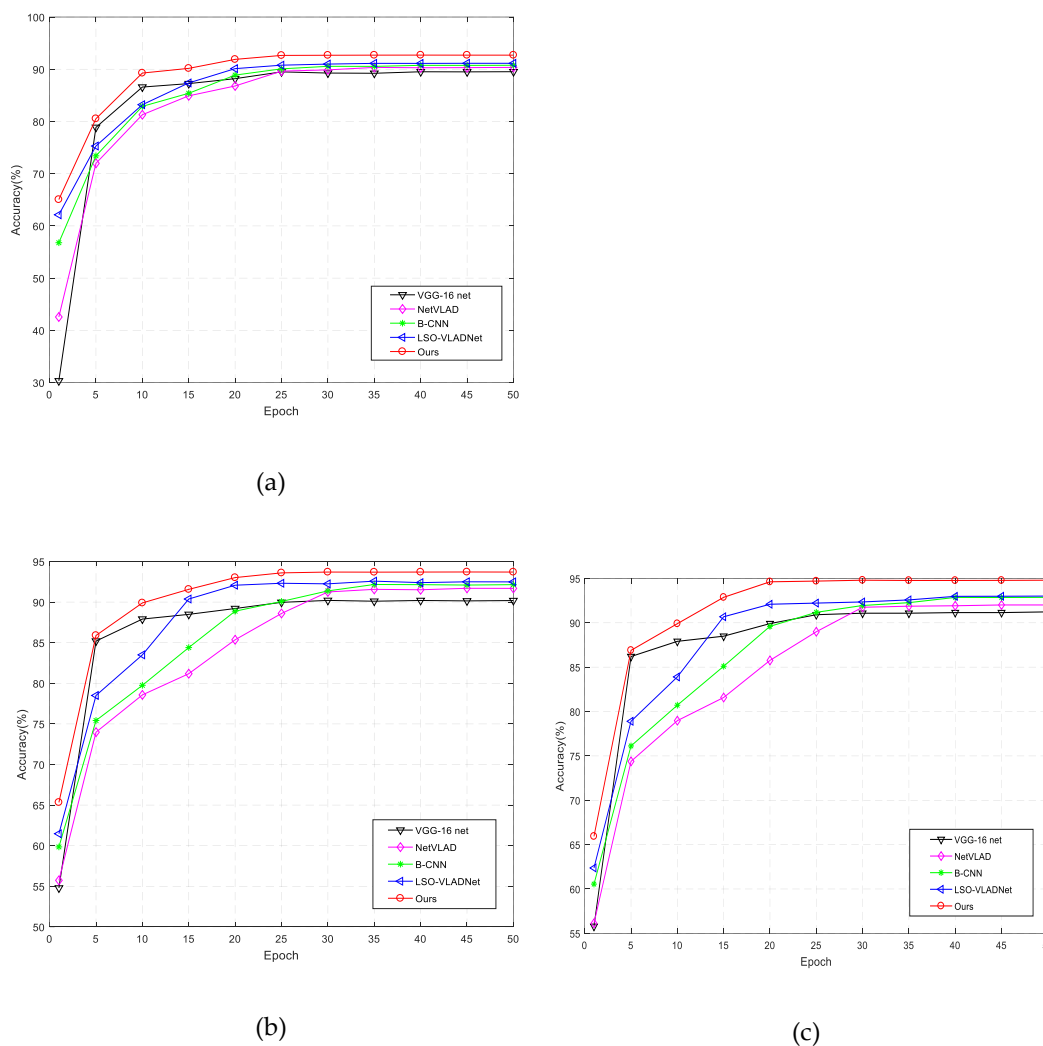


Figure 8. Comparison between the GCCH network and other state-of-the-art methods on the TerraSAR-X dataset. (a) 10%. (b) 20%. (c) 30%.

Table 3. The performance results compared with other state of the art methods.

Method	TerraSAR-X Dataset		
	10%	20%	30%
ResNet-50(fine-tuning) [47]	90.10	91.28	92.23
GoogleNet(fine-tuning) [48]	89.71	90.40	91.74
VGG-16net(fine-tuning) [10]	89.53	90.21	91.23
NetVLAD [37]	90.36	91.75	92.01
B-CNN [44]	90.76	92.35	92.87
LSO-VLADNet [39]	91.14	92.58	92.91
Ours	92.70	93.70	94.80

Because the competing methods are all based on VGG-16 net, Figure 8 only shows the comparison of VGG-16 net. As displayed in Table 3 and Figure 8, NetVLAD is a first-order generalized VLAD layer network (the feature dimension is 512), and achieves better performance than VGG-16 net (fine-tuning the last FC layer). LSO-VLADNet is a localized and second-order VLAD network and extends the novel feature coding approach to the end-to-end model (the feature dimension is $512 \times 2 = 1024$), and therefore it exceeds the performance of NetVLAD and VGG-16 net. The B-CNN method could acquire global feature representation using the outer product at each location of the images (we also can see it as the

first-order network in our paper) and models local pairwise feature interactions in a translationally invariant manner (the feature dimension is $512 \times 512 = 262144$); B-CNN achieves similar performance as LSO-VLADNet, but it only considers the first-order feature representation and high dimension. The GCCH network enables low-dimensional but highly discriminative feature descriptors, which can learn the high-order of parameterized locality-constrained affine subspace coding by the kernelled outer product automatically, and the squeeze and excitation block learns the channel information of each feature representation and helps to boost feature discriminability. Therefore, the proposed GCCH network has advantages (both convergence and accuracy) over these state-of-art methods on the TerraSAR-X dataset, at about 2%-3% larger than the well-known Resnet-50, Google Net and VGG-16 net method (fine tuning).

4.3.4. Comparison with Other Mid-Level Feature Representation Methods

In this section, Table 4 demonstrates the performance comparisons between the proposed network and other mid-level feature representation methods, which are based on handcrafted features (the scale-invariant feature transform (SIFT) feature is used in this subsection). The experimental parameter settings are set the same as in Section 4.3.3, all of the experiments are run 10 times except the proposed network, the resulting features are classified by the linear support vector machine (SVM), and the overall accuracy (OA) is demonstrated in the following table.

Table 4. Comparison with other mid-level feature representation methods.

Method	TerraSAR-X Dataset		
	10%	20%	30%
Bow [49]+SIFT [50]	80.39 ± 1.03	83.22 ± 0.53	85.06 ± 0.23
SC [21]+SIFT [50]	81.09 ± 0.89	83.50 ± 0.27	86.21 ± 0.31
LLC [22]+SIFT [50]	75.51 ± 0.64	78.33 ± 0.21	79.88 ± 0.38
VLAD [38]+SIFT [50]	84.90 ± 0.82	88.87 ± 0.84	91.25 ± 0.67
FV [24]+SIFT [50]	86.65 ± 0.81	91.13 ± 0.60	92.48 ± 0.42
GCCH(Ours)	92.70	93.70	94.80

Table 4 demonstrates that the larger training ratio is, the better the classification will be in most cases. Sparse coding (SC) is one of the soft-assignment methods and decomposes each feature into a sparse, linear combination of the visual words; therefore, the SC method has better performance than the traditional Bow approach (hard-assignment) in most cases. Different from SC and LLC, VLAD is the first-order feature coding style which most fully considers the redundancy of the local geometric structure around visual words, and therefore the feature representation is more efficient than the zeroth-order feature coding method. The FV method, which is a second-order feature coding style, computes gradients of log-likelihood functions of features with respect to covariance and mean vectors; thus, it could obtain the highest score except the proposed network (ours). The GCCH network learns the high-order of parameterized LASC (second order feature coding) by the kernelled outer product automatically; additionally, the squeeze and excitation block learns the channel information of statistic feature representation. Checking in Table 4, we can observe that the GCCH network outperforms all the other mid-level feature representation methods with an increase between 2% and 6.8%.

The confusion matrix reports the detailed classification results of each scene label; each column represents the predicted class, and each row represents the real class. As displayed in Figure 9, some scene categories such as agricultural, beach and runway cannot be effectively classified in the LLC+SIFT and VLAD+SIFT methods, but most categories are classified correctly and produce a classification accuracy higher than 90% in the proposed GCCH network; especially, the agricultural and intersection attain a classification accuracy higher than 95%. In addition, all of the scene classes get a higher score than the LLC+SIFT and VLAD+SIFT approaches. Therefore, the feature descriptors extracted by our proposed network are more discriminative than others.

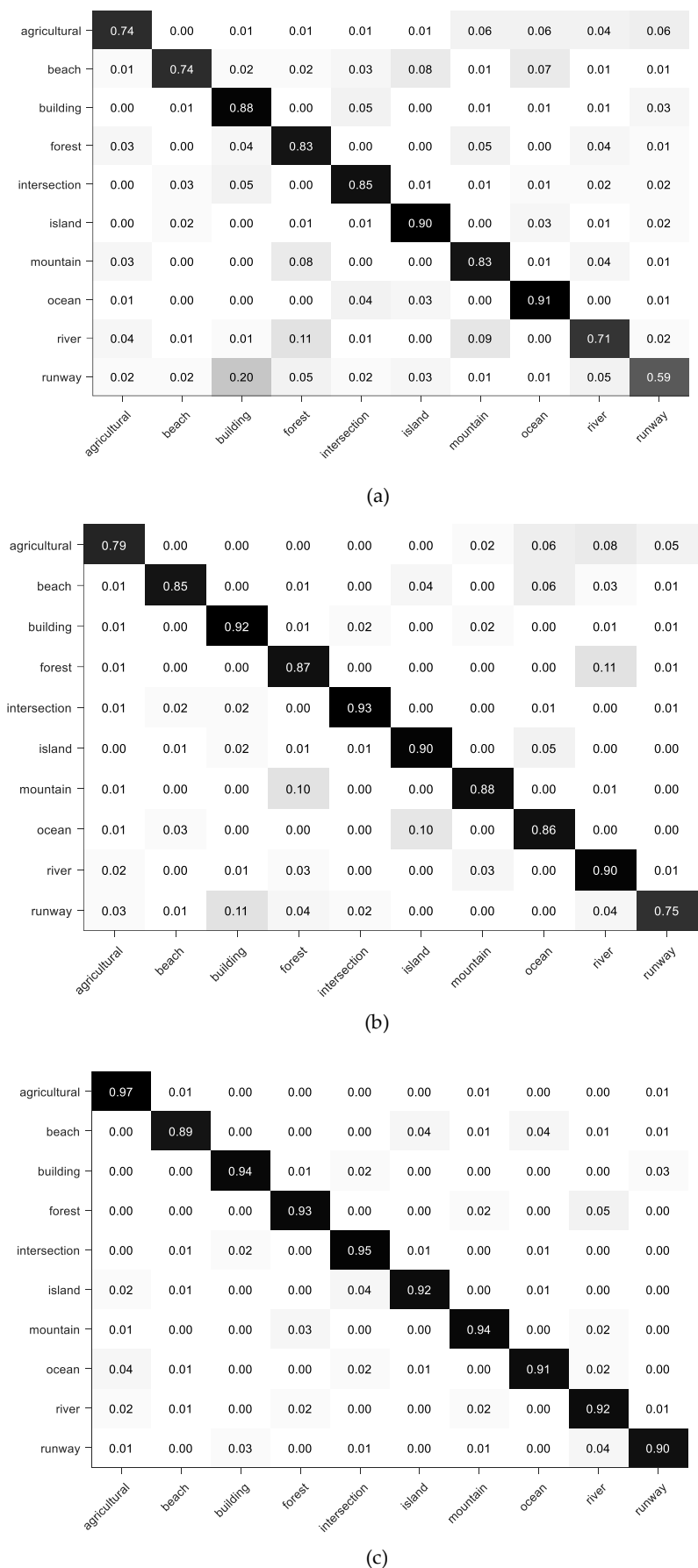


Figure 9. Confusion matrices of the TerraSAR-X dataset under the training ratio of 10% using the following methods: (a) LLC+SIFT. (b) VLAD+SIFT. (c) GCCH (ours).

4.3.5. The Ablation and Combined Experiments

We performed ablation and combined experiments to demonstrate the effectiveness of each module (parameterized LASC layer (PLASC), squeeze and excitation block (SE) and compact high-order generalized orderless pooling layer (CHG)). In this subsection, we truncate the conv5_3 layer of VGG-16 net as the standard convolution layer and make experiments under three different training-test ratios: 10%-90%, 20%-80% and 30%-70%.

We can observe that the VGG-16+CHG approach, which can be viewed as the first-order feature representation, achieves better performance than VGG-16 (fine tuning); meanwhile, LASC is locality-constrained affine subspace coding and constructs a dictionary consisting of an ensemble of linear subspaces attached to representative points, and PLASC contains the first and second-order feature representation. Therefore, vgg-16+PLASC is better than VGG-16 (fine tuning) and VGG-16+CHG. VGG-16+PLASC+SE learns the channel information of the parameterized LASC statistic representation by explicitly modelling the interdependencies between channels and acquires a higher score than VGG-16+PLASC. Similarly, VGG-16+PLASC+CHG can learn generalized compact high-order feature descriptors, which is more discriminative than VGG-16+PLASC and VGG-16+CHG. The GCCH network not only learns the channel information of the parameterized LASC statistic representation, but also acquires the high-order feature descriptors of channel-boosted parameterized LASC statistic representation; therefore, the proposed GCCH network acquires the highest score.

In addition, we introduce the computational complexity between the GCCH network and the ablation and combined networks in Table 5. The SAR remote sensing scene image resolution is 256×256 . The computational efficiency of each module is reported in detail. The proposed GCCH network truncates the conv5_3 layer of VGG-16 net, which does not contain the last three fully connected layers in VGG-16 net, and replaces them with PLASC, SE and CHG (they contain fewer parameters than the last three fully connected layers in VGG-16 net); therefore, the GCCH network runs faster than the VGG-16 (fine tuning) network.

Table 5. The ablation and combined experiments of the GCCH network.

Method	Time (Second)	TerraSAR-X Dataset		
		10%	20%	30%
vgg-16(finertuning)	4.92	89.53	90.21	91.23
vgg-16+PLASC	1.77	91.14	92.58	92.91
vgg-16+CHG	1.53	90.51	92.12	92.38
vgg-16+PLASC+SE	3.58	91.42	92.60	93.12
vgg-16+PLASC+CHG	2.98	91.55	92.65	93.38
GCCH(Ours)	4.14	92.70	93.70	94.80

5. Conclusions

The paper presents a novel effective network for SAR image scene classification, named the generalized compact channel-boosted high-order orderless pooling network (GCCH), which generates compact high-order and low-dimensional but highly discriminative feature descriptors. The proposed network learns the high-order of parameterized LASC (second-order feature coding style) by the kernelled outer product automatically, and the squeeze and excitation block is employed to learn the channel information of statistic feature representation by explicitly modelling the interdependencies between channels. All of the layers can be trained by back-propagation, and the parameters enable end-to-end optimization. Experimental results on the TerraSAR-X dataset demonstrate that the proposed GCCH network further improves the SAR scene classification performance, and performs better than other methods.

The full statistics of convolutional activations and higher-order feature coding style will be considered in future work. Furthermore, the high-order feature statistics information encoding on different network layers is worthy of study.

Author Contributions: K.N. designed and proposed the main structure of this research, analyzed the data and wrote the manuscript. Y.W and P.W. developed this study together. All authors discussed and reviewed the manuscript.

Funding: This research was funded by the Open Project Program of the National Laboratory of Pattern Recognition (NLPR) under Grant 201900029; National Natural Science Foundation of China under Grant No.61573183, No. 61801211.

Acknowledgments: We thank the National Natural Science Foundation of China and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR) for general support of this work. The authors would like to thank the handling editors and the reviewers for providing valuable comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ren, Z.L.; Hou, B.; Wen, Z.D.; Jiao, L.C. Patch-sorted deep feature learning for high resolution SAR image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 3113–3126. [[CrossRef](#)]
2. Esch, T.; Schenk, A.; Ullmann, T.; Thiel, M.; Roth, A.; Dech, S. Characterization of land cover types in terrasar-x images by combined analysis of speckle statistics and intensity information. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1911–1925. [[CrossRef](#)]
3. Kwon, T.J.; Li, J.; Wong, A. Etvos: An enhanced total variation optimization segmentation approach for SAR sea-ice image segmentation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 925–934. [[CrossRef](#)]
4. Yang, X.; Clausi, D. Evaluating SAR sea ice image segmentation using edge-preserving region-based MRFs. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2012**, *5*, 1383–1393. [[CrossRef](#)]
5. Knuth, R.; Thiel, C.; Thiel, C.; Eckardt, R.; Richter, N.; Schmullius, C. Multisensor SAR analysis for forest monitoring in boreal and tropical forest environments. In Proceedings of the IEEE International Geoscience & Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009.
6. Yang, S.; Wang, M.; Feng, Z.; Liu, Z.; Li, R. Deep sparse tensor filtering network for synthetic aperture radar images classification. *IEEE T Neural Netw. Learn. Syst.* **2018**, *29*, 3919–3924. [[CrossRef](#)]
7. Yang, S.; Wang, M.; Long, H.; Liu, Z. Sparse robust filters for scene classification of synthetic aperture radar (SAR) images. *Neurocomputing* **2006**, *184*, 91–98. [[CrossRef](#)]
8. Geng, J.; Wang, H.; Fan, J.; Ma, X. SAR image classification via deep recurrent encoding neural networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 2255–2269. [[CrossRef](#)]
9. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
11. Li, E.Z.; Xia, J.S.; Du, P.J.; Lin, C.; Samat, A. Integrating multilayer features of convolutional neural networks for remote sensing scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5653–5665. [[CrossRef](#)]
12. He, N.J.; Fang, L.Y.; Li, S.T.; Plaza, A.; Plaza, J. Remote sensing scene classification using multilayer stacked covariance pooling. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6899–6910. [[CrossRef](#)]
13. Matthew, D.Z.; Rob, F. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2014; pp. 818–833.
14. Gong, Y.C.; Wang, L.W.; Guo, R.Q.; Lazebnik, S. Multi-scale orderless pooling of deep convolutional activation features. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2014; pp. 392–407.
15. Silva, F.B.; Werneck, R.D.O.; Goldenstein, S.; Tabbone, S.A. Graph-based bag-of-words for classification. *Pattern Recogn.* **2017**, *74*, 266–285. [[CrossRef](#)]
16. Xie, X.M.; Zhang, Y.Z.; Wu, J.J.; Shi, G.M.; Dong, W.S. Bag-of-words feature representation for blind image quality assessment with local quantized pattern. *Neurocomputing* **2017**, *266*, 176–187. [[CrossRef](#)]
17. Sun, H.; Sun, X.; Wang, H.; Li, Y.; Li, X. Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 109–113. [[CrossRef](#)]
18. Li, Y.; Liu, L.Q.; Shen, C.H.; Hengel, A.V.D. Mining mid-level visual patterns with deep CNN activations. *Int. J. Comput. Vis.* **2015**, *121*, 971–980. [[CrossRef](#)]
19. Sivic, J. A text retrieval approach to object matching in videos. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003.

20. Lazebnik, S. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006.
21. Sheng, G.F.; Yang, W.; Xu, T.; Sun, H. High-resolution satellite scene classification using a sparse coding based multiple feature combination. *Int. J. Remote Sens.* **2012**, *33*, 2395–2412. [[CrossRef](#)]
22. Wang, J.; Yang, J.; Kai, Y.; Lv, F.; Huang, T.; Gong, Y. Locality-constrained Linear Coding for image classification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010.
23. Zhou, X.; Yu, K.; Zhang, T.; Huang, T.S. Image classification using super-vector coding of local image descriptors. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2010; Volume 6315, pp. 141–154.
24. Sánchez, J.; Perronnin, F.; Mensink, T.; Verbeek, J. Image classification with the fisher vector: Theory and practice. *Int. J. Comput. Vis.* **2013**, *105*, 222–245. [[CrossRef](#)]
25. Li, P.; Lu, X.; Wang, Q. From dictionary of visual words to subspaces: Locality-constrained affine subspace coding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
26. Zhu, Q.; Zhong, Y.; Zhao, B.; Xia, G.S.; Zhang, L. Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 747–751. [[CrossRef](#)]
27. Hu, F.; Xia, G.S.; Hu, J.W.; Zhang, L.P. Transferring deep Convolutional Neural Networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [[CrossRef](#)]
28. Qi, K.; Wu, H.; Shen, C.; Gong, J. Land-use scene classification in high-resolution remote sensing images using improved correlators. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2403–2407.
29. Bian, X.; Chen, C.; Tian, L.; Du, Q. Fusing local and global features for high-resolution scene classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 2889–2901. [[CrossRef](#)]
30. Mekhalfi, M.L.; Melgani, F.; Bazi, Y.; Alajian, N. Land-use classification with compressive sensing multifeature fusion. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2155–2159. [[CrossRef](#)]
31. Cheng, G.; Li, Z.; Yao, X.; Guo, L.; Wei, Z. Remote sensing image scene classification using bag of convolutional features. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1735–1739. [[CrossRef](#)]
32. Yuan, B.; Li, S.; Li, N. Multiscale deep features learning for land-use scene recognition. *J. Appl. Remote Sens.* **2018**, *12*, 015010. [[CrossRef](#)]
33. Banerjee, B.; Chaudhuri, S. Scene recognition from optical remote sensing images using mid-level deep feature mining. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1080–1084. [[CrossRef](#)]
34. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th Sigspatial International Conference on Advances in Geographic Information Systems, New York, NY, USA, 2–5 November 2010.
35. Fan, J.; Chen, T.; Lu, S. Unsupervised feature learning for land-use scene recognition. *IEEE Trans. Geosci. Remote Sens.* **2017**, *50*, 2250–2261. [[CrossRef](#)]
36. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005.
37. Arandjelovic, R.; Gronat, P.; Torii, A.; Pajdla, T.; Sivic, J. Netvlad: CNN architecture for weakly supervised place recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1437–1451. [[CrossRef](#)]
38. Jegou, H.; Douze, M.; Schmid, C.; Perez, P. Aggregating local descriptors into a compact image representation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010.
39. Chen, B.; Li, J.; Wei, G.; Ma, B. A novel localized and second order feature coding network for image recognition. *Pattern Recogn.* **2018**, *76*, 339–348. [[CrossRef](#)]
40. Ni, K.; Wang, P.; Wu, Y.Q. High-order generalized orderless pooling networks for synthetic-aperture radar scene classification. Available online: http://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html (accessed on 23 April 2019).
41. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. Available online: <https://ieeexplore.ieee.org/abstract/document/8695749> (accessed on 29 April 2019).

42. Nogueira, K.; Penatti, O.A.B.; Santos, J.A.D. Towards better exploiting Convolutional Neural Networks for remote sensing scene classification. *Pattern Recogn.* **2017**, *61*, 539–556. [[CrossRef](#)]
43. Lin, R.; Xiao, J.; Fan, J. Nextvlad: An efficient neural network to aggregate frame-level features for large-scale video classification. In Proceedings of the Workshop on Statistical Learning in Computer Vision ECCV, Munich, Germany, 8–14 September 2018; pp. 206–218.
44. Lin, T.Y.; Roychowdhury, A.; Maji, S. Bilinear cnn models for fine-grained visual recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1449–1457.
45. Kong, S.; Fowlkes, C. Low-rank bilinear pooling for fine-grained classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
46. Gao, Y.; Bejbom, O.; Zhang, N.; Darrell, T. Compact bilinear pooling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 317–326.
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
48. Zegey, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
49. Sun, Y.; Liu, Z.P.; Todorovic, S.; Li, J. Adaptive boosting for SAR automatic target recognition. *IEEE T Aero Elec Sys.* **2005**, *43*, 112–125. [[CrossRef](#)]
50. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).