remote sensing

MDPI

Article

Color-Boosted Saliency-Guided Rotation Invariant Bag of Visual Words Representation with Parameter Transfer for Cross-Domain Scene-Level Classification

Li Yan, Ruixi Zhu *, Yi Liu * and Nan Mo

School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; lyan@sgg.whu.edu.cn (L.Y.); nmo@whu.edu.cn (N.M.)

* Correspondence: ruixzhu@whu.edu.cn (R.Z.); yliu@sgg.whu.edu.cn (Y.L.)

Received: 21 March 2018; Accepted: 12 April 2018; Published: 15 April 2018

Abstract: Scene classification on remote sensing imagery is usually based on supervised learning but collecting labelled data in remote sensing domains is expensive and time-consuming. Bag of Visual Words (BOVW) achieves great success in scene classification but there exist problems in domain adaptation tasks, such as the influence of background and the rotation transformation on BOVW representation, and the transfer of SVM parameters from the source domain to the target domain, which may lead to decreased cross-domain scene classification performance. In order to solve the three problems, Color-boosted saliency-guided rotation invariant bag of visual words representation with parameter transfer is proposed for cross-domain scene classification. The global contrast-based salient region detection method is combined with the color-boosted method to increase the accuracy of detected salient regions and reduce the effect of background information on the BOVW representation. The rotation invariant BOVW representation is also proposed by sorting the BOVW representation in each patch in order to decrease the effect of rotation transformation. The several best configurations in the source domain are also applied to the target domain so as to reduce the distribution bias between scenes in the source and target domain. These configurations deliver the top classification performance the optimal parameter in the target domain. The experimental results on two benchmark datasets confirm that the proposed method outperforms most previous methods in scene classification when instances in the target domain are limited. It is also proved that color boosted global contrast-based salient region detection (CBGCSRD) method, rotation invariant BOVW representation, and transfer of SVM parameters from the source to the target domain are all effective in improving the classification accuracy with 2.5%, 3.3%, and 3.1%. These three contributions may increase about 7.5% classification accuracy in total.

Keywords: Bag of Visual Words (BOVW); color-boosted global contrast-based salient region detection method; rotation invariant BOVW representation; transfer of SVM parameters from the source to the target domain; cross-domain scene classification

1. Introduction

With the development of remote sensing sensors, satellite image sensors can offer images with a spatial resolution at the decimeter level. We call these images high-resolution remote sensing images (HRIs). However, despite enhanced resolution, these details often suffer from the spectral uncertainty problems stemming from an increase of the intra-class variance [1] and a decrease of the inter-class variance [2]. Taking into account these characteristics, HRIs classification methods have evolved from pixel-oriented methods to object-oriented methods and have achieved precise object recognition performance [3–5]. However, these methods may lead to the so-called "semantic gap" [6], namely the

divergence between low-level data and high-level semantic information. In order to better acquire the semantic information in accordance with human cognition, scene classification aimed at automatically labeling an image from a set of semantic categories [7], has been proposed with remarkable success in image interpretation.

Scene classification with HRI is usually based on supervised learning method requiring a set of new collected samples [8,9]. However, we are usually faced with the tasks that the target images are with a limited number of samples, but we have sufficient previous labeled data. In many cases, previous labeled data remains useful for training a new target classifier [10,11]. However, the direct application of previous instances to new remote sensing images often provides poor results because the spectra observed in the new scene is highly different from that in the existing scenes even though they represent the same types of objects. This is due to a variety of factors, such as changes in the acquisition conditions including the illumination or acquisition angle, seasonal changes, or the use of different sensors. This problem can be suppressed by transfer learning methods [12].

As mentioned by Devis Tuia [13], transfer learning methods in the remote sensing literature can be categorized into four categories:

(1) Selection of invariant features. These methods are usually achieved by considering only a subset of the original features that are invariant between the domains [14–18].
(2) Adaptation of the data distributions. Data is adapted such that the feature distributions over the different domains are more compatible. This technique is also known as feature extraction and representation learning [19–25].
(3) Adaptation of the classifier with semi-supervised method. Techniques belonging to this family take a semi-supervised strategy that utilizes the unlabeled target samples to adapt a classifier trained using the labeled source samples [26–31].
(4) Adaptation of the classifier by active learning (AL). These techniques also utilize a semi-supervised strategy. However, instead of automatically labeling samples from the target domain, these techniques require the user to label some target samples. Therefore, the main challenge here is how to select the minimal set of informative target samples that the expert user needs to annotate [32–36].

In order to recognize and analyze scenes from remote sensing images, various scene classification methods have been proposed over decades. One particular method called the bag-of-visual-words (BOVW) [37,38], has been successfully utilized for scene classification. The BOVW approach treats an image as a collection of unordered feature descriptors, and represents images with the frequency of "visual words" that are constructed by quantizing local features. It can be divided into two parts: dictionary learning and feature encoding. Dictionary learning consists of clustering feature descriptors in each local patch and using the resulting clusters as visual vocabularies. These visual vocabularies can be used for the feature encoding. In the feature encoding step, the images are finally represented by the unordered collections of the visual vocabularies and the histograms of the occurrences concerning the visual vocabularies.

However, the BOVW representation does not perform well in the adaptation of the data distributions due to following reasons:

(1) The influence of background information in images. Existing BOVW representations mainly extract features from the whole image rather than the salient regions, which may lead to higher difference between representations from both domains owing to more complex scenes in remote sensing domains.
(2) The effect of rotation transformation. Existing BOVW methods suffer from rotation transformations since patches in spatial pyramid matching (SPM) [39] are in a fixed order. Therefore, when faced with more serious rotation transformations in images from both domains, poor cross-domain scene classification performance may be delivered due to aggravated feature bias between two domains.

(3) The transfer of SVM parameters from the source to the target domain. The performance of support vector machine classifiers [40] is directly influenced by the values of its free parameters. The free parameters of instances in the source domain are different from those in the target domain because of different feature distributions. Therefore, more optimal parameters for the target domain need to be adjusted to the free parameters in the source domain.

The BOVW representation considers feature representations of a set of patches from SPM in the whole image. However, it may be affected by noisy information, such as the background or other irrelevant objects, which may increase the feature difference in the representations of the same category from different domains. In order to solve this problem, various studies have been conducted on salient region detection to detect the salient objects in the same category from both domains for higher similarity in BOVW representations, such as airplanes in airports, cars in parking lots, and so on. The salient region detection methods can be categorized into two types: local contrast based methods and global contrast methods [41].

Local contrast based methods investigate the rarity of image regions with respect to local neighborhoods. Itti [42] defines image saliency using central-surrounded differences between multi-scale image features based on the highly influential biologically-inspired early representation model introduced by Koch and Ullman [43]. Ma and Zhang [44] propose an alternate local contrast analysis for generating saliency maps, which is then extended using a fuzzy growth model. Liu [45] finds multi-scale contrast by linearly combining contrast to a Gaussian image pyramid. More recently, Goferman [46] simultaneously modeled local low-level clues, global considerations, visual organization rules, and high-level features to highlight salient objects along with their contexts. Such methods using local contrasts tend to produce higher saliency values near edges instead of uniformly highlighting salient regions in images, which may remove regions corresponding to relevant objects.

In order to solve the problem of the local contrast based methods, global contrast-based methods have been proposed over decades. Global contrast-based methods evaluate saliency of an image region using its contrast with respect to the entire image. Zhai and Shah [47] define pixel-level saliency based on a pixel's contrast to all other pixels. However, for efficiency they use only luminance information, thus ignoring distinctiveness clues in other channels. Achanta [48] proposes a frequency tuned method that directly defines pixel saliency using a pixel's color difference from the average image color. The elegant approach, however, only considers first order average color, which can be insufficient to analyze complex variations in remote sensing images. Ko and Nam [49] select salient regions using a support vector machine trained on image segment features, and then cluster these regions to extract salient objects. Han et al. [50] model color, texture, and edge features of a Markov random field framework to grow salient object regions from seed values in the saliency maps. Cheng [41] proposes the saliency of one region depends on its contrast with respect to its nearby regions, while contrasts to distant regions are less significant. However, when faced with complex scenes such as remote sensing images, the global contrast based saliency detection methods usually deliver poor performance because of relatively low contrast between background and salient objects, which may misclassify the background regions as salient regions [51]. Therefore, a color-boosted method originally in salient point detectors [52] has been introduced to global contrast-based salient region detection to increase the color contrast between different regions in images for more accurate salient region detection.

BOVW with the SPM method mostly uses ordered regular grid or block partition of an image to incorporate spatial information [53] and, therefore, are sensitive to the rotation transformation of scenes, which will inevitably result in misclassification of scene images that belong to the same category and influence the classification accuracy. As shown in Figure 1, many approaches based on the concentric circle-based partition strategy of an image in color and texture feature extraction has been proposed. Qi [54] proposes a multi-scale deeply-described correlatons (MDDC)-based model by applying adaptive vector quantization of multi-scale correlograms to achieve rotation invariant representation without loss of discrimination. Zhao [55] proposes a concentric circle-structured multi-scale BOVW model by partitioning the image into a series of annular sub-regions and computing

histograms of local features found inside each annular sub region. Khan [56] proposes a pairwise spatial histogram by concentric circles and angles centered on each descriptor. However, the spatial layout information may be lost to some degree since we only know these patches are adjacent but we do not know the specific spatial relationships. To solve this problem, a shifting operation is proposed to BOVW representation with an order of ascending distance between zero vector and feature vector in one sub-image from eight sub-images adjacent to each other uniformly segmented from the whole image, which is a simple, but effective, way to incorporate rotation-invariant spatial layout information of scene images into the original BOVW model.

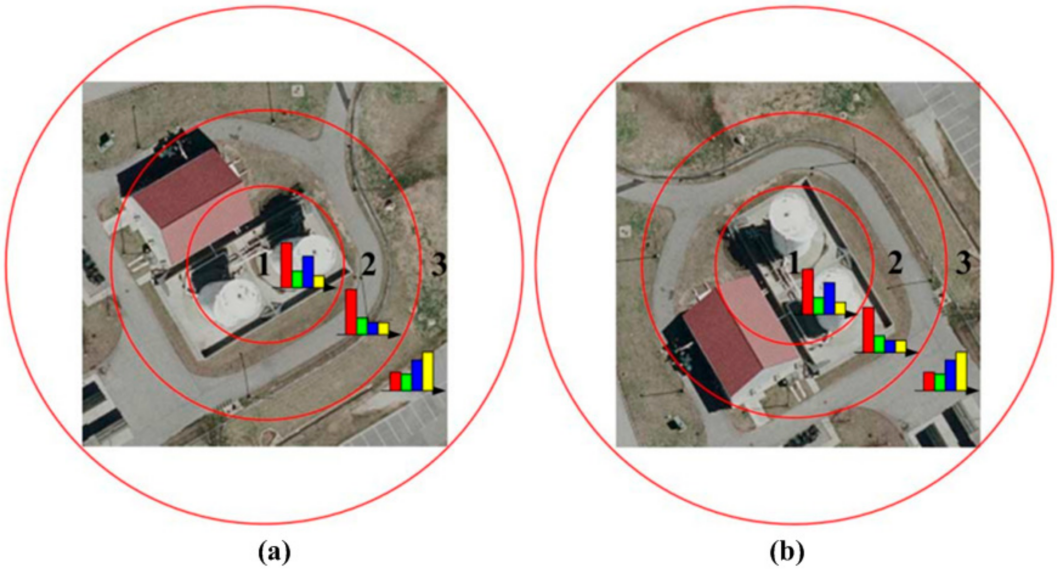


Figure 1. Example of spatial partition of rotated remote sensing images using a concentric circle-based partition strategy; (**a**) Original image, and (**b**) rotated image.

The BOVW representations will be put into a support vector machine (SVM) classifier for cross-domain scene classification. The performance of SVM is directly influenced by the choice of kernel function and values of its free parameters, such as the penalty for the cost function c and the coefficient for the kernel function g. However, the parameters of SVM classifiers from instances in the source domain need to be adjusted to those in the target domain due to feature distribution bias between instances in both domains. Regarding the transfer of parameters, most approaches assume that the individual SVM for different, but related, domains must share some parameters featuring a transfer supervised learning. Some studies closely related to transfer onto SVM parameters are described next.

Soares et al. [57] propose a meta-learning methodology that explores information about the past performance of different parameters. The methodology is applied to adjust the width of the Gaussian kernel for regression problems with low error while providing significant savings in time. In [58], particle swarm optimization (PSO) was applied to the problem of parameter tuning of support vector machines. As learning systems are essentially multi-objective problems, the multi-objective PSO was used to maximize the success rate and minimize the complexity of the model with faster search process convergence speed and less computational cost. Ideas of meta-learning and case-based reasoning have been used to provide good starting points for genetic algorithms to find good parameters for support vector machines and random forests. The presented approach achieves accuracy comparable to grid search with a significantly lower computational cost. Reif [59] uses ideas of meta-learning and case-based reasoning to provide good starting points for genetic algorithm to find good parameters for SVM classifiers. Ali and Smith-Miles [60] used goal-based learning rules for automatic kernel selection with empirical evaluation based on classification.

However, the aforementioned studies focus on the search techniques or the choice of kernel functions may suffer from heavy computational burden due to too large a search space or enormous manual interventions on labeling instances in the target domain. In order to suppress the problems, the several top pre-calculated parameter configurations organized in decreasing order of classification performance in the source domain are proposed to be transferred to the target domain and the parameter with the best classification performance in the target domain will be the optimal parameter to reduce the computational cost of parameter transfers and increase classification accuracy in the target domain with limited samples in the target domain.

Inspired by the aforementioned work, we propose a color-boosted saliency-guided rotation invariant BOVW model with a transfer of SVM parameters from the source domain to the target domain. The main contributions of this paper are summarized below:

(1) A color-boosted method has been introduced to global contrast based salient region detection method to increase the color contrast between different regions and obtain salient regions for BOVW representations in order to reduce the effect of background or non-salient objects on the BOVW representation.
(2) A shifting operation has been applied to represent the images with BOVW representations of patches in an order of ascending distance from zero to feature vectors in the sub-image rather than those of patches in SPM so as to decrease the effect of rotation transformation on classification accuracy.
(3) Several pre-calculated best parameters in the source domain have been transferred to the target domain in a decreasing order and the parameter with the best performance in the target domain will be the optimal parameter setting to reduce the required number of instances in the target domain and the effort in searching for the optimal parameters.

The rest of the paper is organized as follows: In Section 2, we describe the overall process and details of the proposed color boosted saliency-guided rotation invariant BOVW approach with parameter transfer. In Section 3, several experiments and results in two benchmark datasets are presented to demonstrate the effectiveness and superiority of the proposed algorithms. In Section 4, a discussion about the proposed method with a parameter sensitivity analysis is conducted. Conclusions and suggestions for future work are summarized in Section 5.

2. Materials and Methods

In this section, we present a cross-domain scene classification method of HRIs based on color-boosted saliency-guided rotation invariant BOVW representation with parameter transfers, as shown in Figure 2, which can be divided into four main steps:

(1) For one category in the source domain, the CBGCSRD method has been applied to calculate the salient region for each instance. Then DenseSIFT descriptors [39] will be only calculated in the salient regions. All extracted descriptors in one category are clustered by K-means to form the codebook for this category.
(2) The instance is segmented into eight regions with the same area. Then, for each region, BOVW representation has been calculated with the codebook in step (1). A shifting operation with patches in an order of ascending Euclidean distance from zero to feature vectors in all regions is used to obtain rotation invariant BOVW representation for instances in the source and target domain from the above codebook to reduce the effect of rotation transformation.
(3) The rotation invariant BOVW representations of instances from the source domain are put into SVM classifiers with several predefined parameters to obtain classification accuracy of instances from the source domain with a decreasing order. Several parameters with the top classification performance in the source domain will be tested for the instances in the target domain.
(4) The several top parameters in the source domain performing best in the instances of target domain will be selected as the optimal parameters and applied to testing images from the target domain.

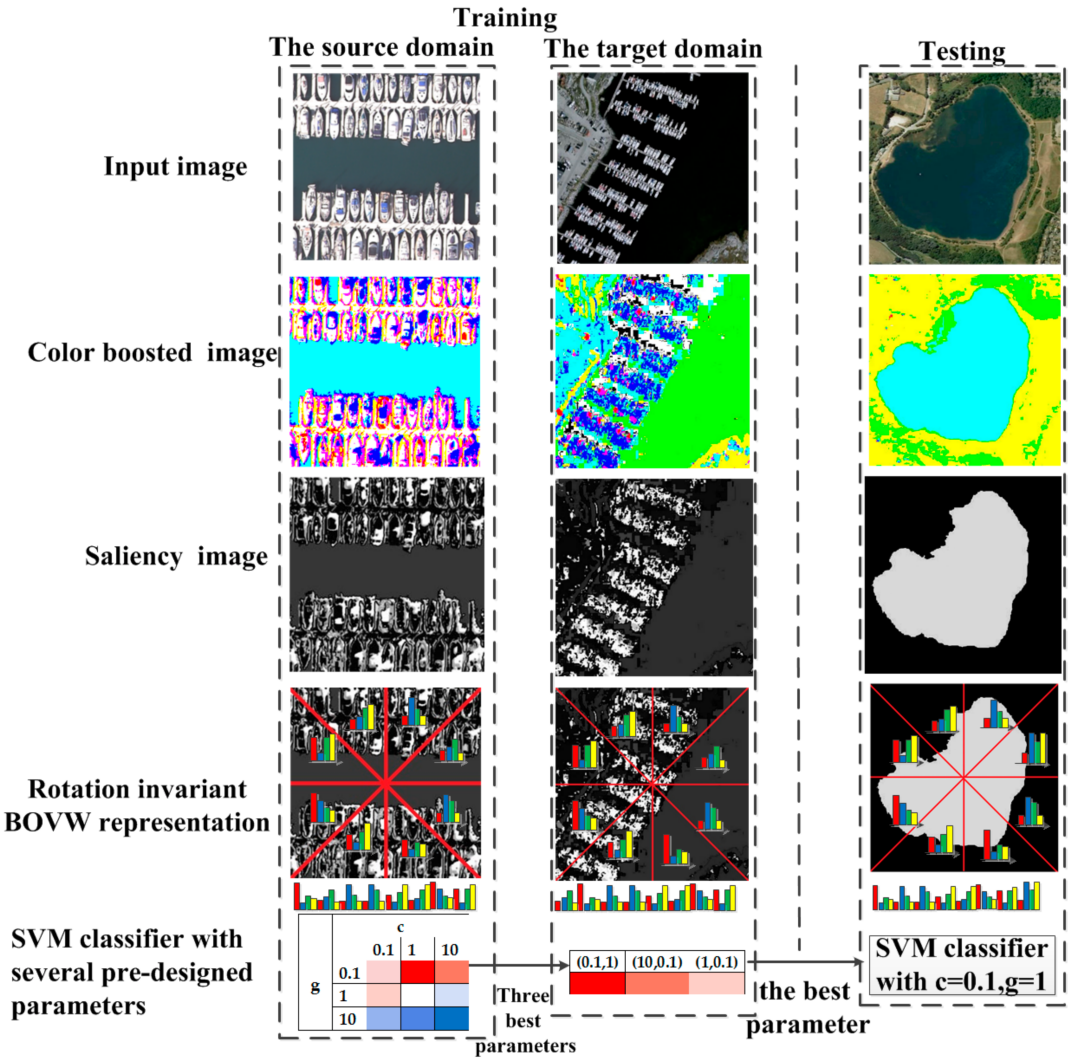


Figure 2. The architecture of the proposed color boosted saliency-guided rotation invariant BOVW approach with parameter transfer. The color indicates the performance of the classifier from dark red to dark blue where the parameter with dark red performs the best and that with dark blue delivers the poorest performance.

2.1. CBGCSRD Method

Color saliency boosting based on an analysis of the statistics of color image derivatives incorporates color distinctiveness into existing salient feature detectors to improve acquired information contents by salient features. Similarly, color saliency boosting can also be used in the saliency detection of remote sensing images for obtaining more important information contents in scenes. In order to solve the problem of poor performance of complex scenes in remote sensing images with global contrast based salient region detection methods, color saliency boosting has been introduced to the regional contrast based saliency detection method.

The channels of (R, G, B) space are correlated due to the physics of the world. Photometric events in the real-world, such as shading, shadows, and specularities, influence (R, G, B) values in a well-defined manner [61]. Therefore, the (R, G, B) space needs to be transformed into a color space which is uncorrelated with respect to these photometric events. (o_1, o_2, o_3) space has been used in this paper since it has been proved effective in feature extraction and object detection.

For each image of (R, G, B) space, a color coordinate transformation has been made for adapting the saliency map to focus on rare color derivatives based on Equation (1):

$$\begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix} = \begin{bmatrix} \alpha & & \\ & \beta & \\ & & \gamma \end{bmatrix} \begin{pmatrix} (R-G)/\sqrt{2} \\ (R+G-2B)/\sqrt{6} \\ (R+G+B)/\sqrt{3} \end{pmatrix}; \quad (1)$$

where $\alpha^2 + \beta^2 + \gamma^2 = 1$. Then the color statistics o_1, o_2 will be multiplied by a factor k to increase color contrast in this color space to obtain ko_1, ko_2, then the color statistics will be transformed back into the (R, G, B) space with Equation (2):

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} ko_1/\sqrt{2} + ko_2/\sqrt{6} + o_3/\sqrt{3} \\ -ko_1/\sqrt{2} + ko_2/\sqrt{6} + o_3/\sqrt{3} \\ -2ko_2/\sqrt{6} + o_3/\sqrt{3} \end{pmatrix}; \quad (2)$$

Then the saliency map will be calculated for color boosted images. First of all, we first segment the input image into regions using a graph-based image segmentation method. The saliency of k-th region r_k can be calculated with Equation (3):

$$S(r_k) = w_s(r_k) \sum_{r_k \neq r_i} e^{\frac{D_s(r_k, r_i)}{-\sigma_s^2}} w(r_i) D_r(r_k, r_i); \quad (3)$$

where $D_s(r_k, r_i)$ reflects the spatial distance between region r_k and r_i, and the spatial distance between two regions is defined as the Euclidean distance between their centroids $w(r_i)$, the weight of the region is r_i, and $D_r(r_k, r_i)$ is the color Euclidean distance between region r_k and r_i. We weight the distances by the number of pixels in r_i, namely $w(r_i)$, to emphasize color contrast to larger regions. For σ_s, larger values of σ_s reduce the effect of spatial weighting so that contrast to farther regions would contribute more to the saliency of the current region and $w_s(r_k)$ is a spatial prior weighting term similar to center bias defined by Equation (4):

$$w_s(r_k) = \exp^{-9{d_k}^2}; \quad (4)$$

where d_k is the average distance between pixels in region r_k and the center of the image. The color distance $D_r(r_1, r_2)$ between two regions r_1 and r_2 is shown in Equation (5):

$$D_r(r, r_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_1, i) f(c_2, j) D(c_{1,i}, c_{2,j}); \quad (5)$$

where $f(c_k, i)$ is the probability of the i-th color $c_{k,i}$ among all n_k in the k-th region and $D(c_{1,i}, c_{2,j})$ reflects the distance between the i-th color in the region r_1 and the j-th color in the region r_2.

As can be seen in Figure 3, when faced with complex scenes, the regional contrast based saliency detection method may detect some non-salient regions due to low contrast between background pixels and salient pixels. The color-boosted map increases the difference between salient and non-salient regions and reduce the possibility of detecting non-salient regions. Only features located in the salient regions are used for the BOVW representation.

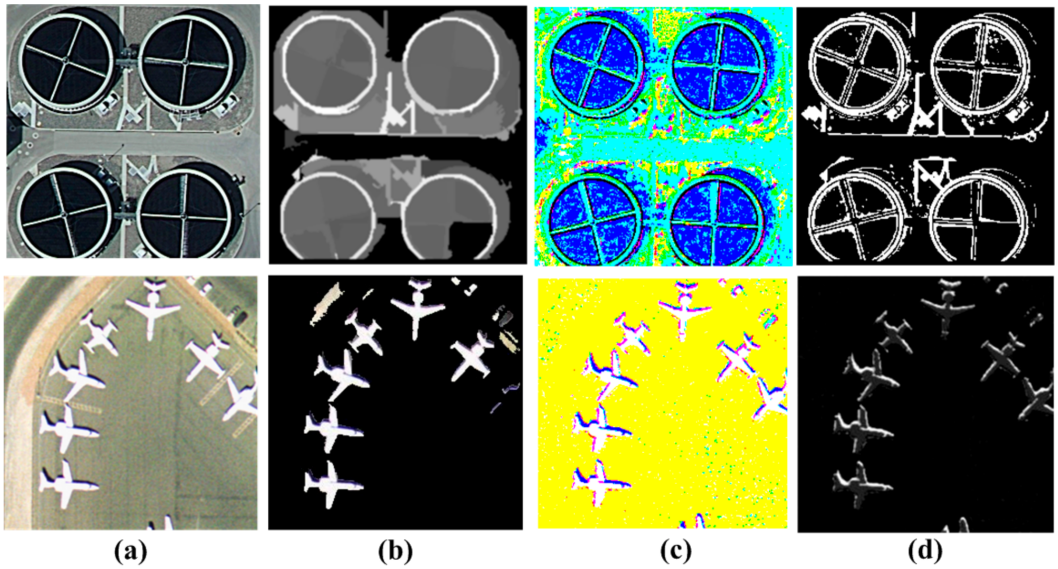


Figure 3. Example of the CBGCSRD method. (**a**) Original image; (**b**) saliency map with existing regional contrast based method; (**c**) color-boosted image; and (**d**) the proposed color-boosted saliency map.

2.2. Rotation Invariant BOVW Representation

BOVW with the SPM method mostly uses ordered regular grid or block partition of an image to incorporate spatial information when generating BOVW representations and, therefore, are sensitive to the rotation transformation of image scenes, which will inevitably result in misclassification of scene images that belong to the same category and influence the classification accuracy, as shown in Figure 4a,b.

When the traditional ordered regular block partitions strategy is used to construct the histogram of visual word occurrences within each subregion, the final orderly concatenated histograms with SPM generated from all subregions will be quite different in the same image but, with rotation, giving rise to misclassification of two similar scene images, which can be seen in Figure 4c.

To achieve rotation-invariance, this paper proposes a rotation invariant BOVW representation to cross-domain scene classification by changing the order of patches in the SPM method to reduce the effect of rotation transformation, which is a simple, but effective, way to incorporate rotation-invariant spatial layout information of scene images into the original BOVW model, as can be seen in Figure 4d.

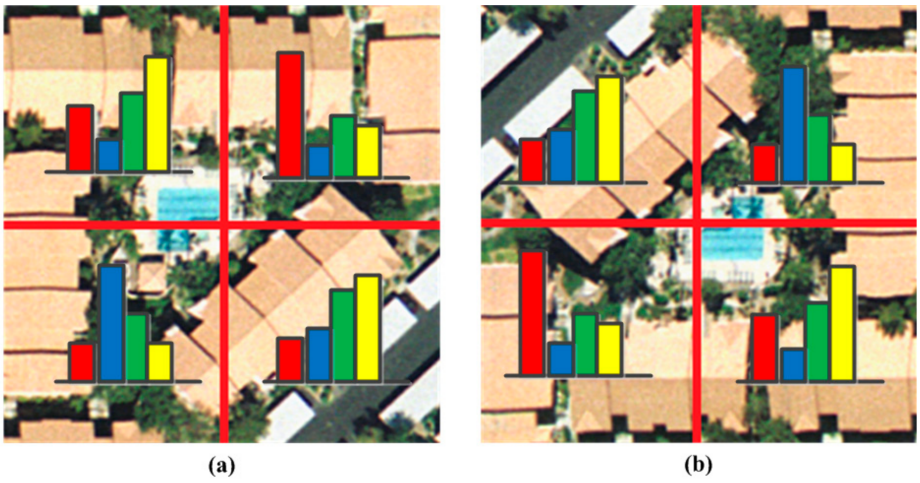


Figure 4. *Cont.*

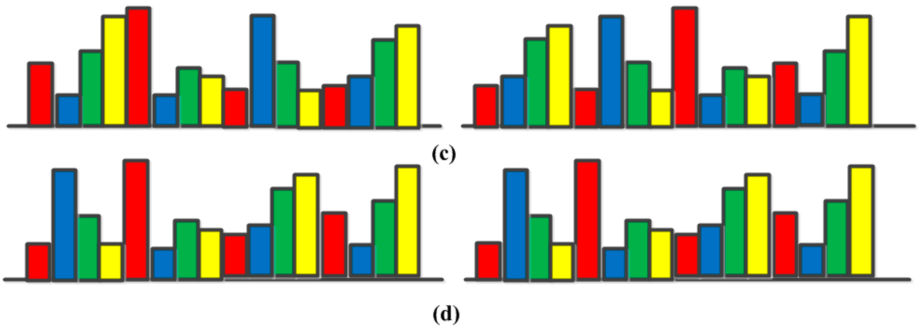


Figure 4. The rotation invariant BOVW representation against typical BOVW representation method on the rotation variance of an image. (**a**) BOVW representations in each patch of original image; (**b**) BOVW representations in each patch of rotated image; (**c**) BOVW representation with SPM for two images; and (**d**) the rotation invariant BOVW representation for two images.

The process of generating rotation invariant BOVW representation is shown in Figure 5, which can be divided into four steps.

(1) Extracting all DenseSIFT features in the color boosted salient regions in one image. All features in one category are clustered by K-means to form the codebook for this category.

(2) In order to represent the image, BOVW feature in each region f_i has been calculated and i is the index of region. While doing so, find the Euclidean distance vectors from zero vectors to feature vector f_i as shown in Equation (6):

$$dist_i = dist(f_i, 0); \tag{6}$$

(3) If $dist_i \geq dist_{i+1}$, then shift the f_i up. Repeat step (2) for all regions with all regions in the order of ascending $dist_i$ for the full image.

(4) Concatenating the BOVW representations with the order in step (3) to form the final rotation invariant BOVW representation.

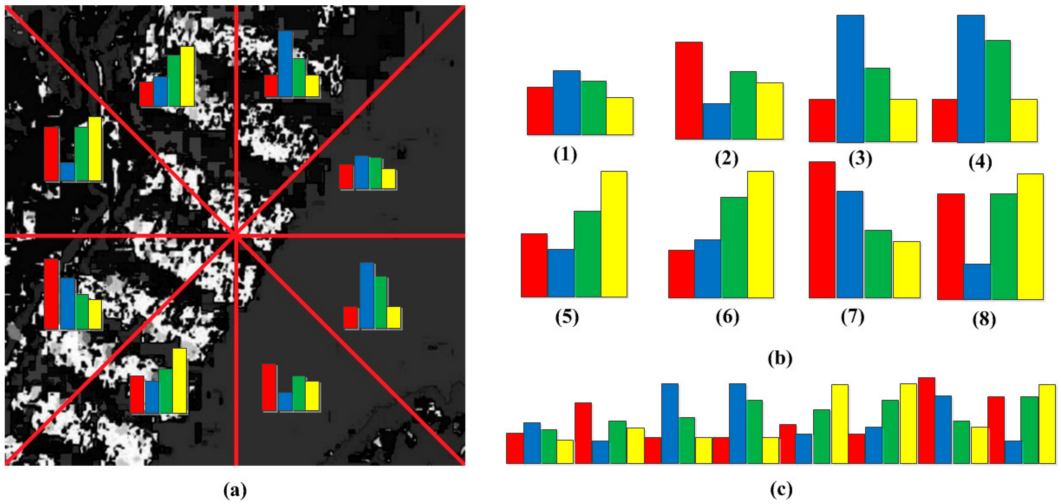


Figure 5. A graphical representation of Rotation invariant BOVW representation. (**a**) The partitions of images in rotation invariant BOVW representation; (**b**) BOVW representation in an ascending order of distance between zero vector and feature vector in each patch; and (**c**) the final rotation invariant BOVW representation.

The rotation invariant BOVW representation is aimed at reducing the effect of rotation transformation on the classification performance.

2.3. Transferring SVM Parameters from the Source Domain to the Target Domain

Above all, instances in the target domain and those in the source domain belong to the same category, ensuring the similarity between instances in both domains since transfer learning methods are based on assumptions that instances in the source domain and those in the target domain are drawn from different, but related, distributions. The proposed approach transfers several best parameter settings in terms of classification accuracy from the source domain to the target domain in order to reduce the search space for parameter transfer to the target domain.

The sketch map of the parameter transfer with nine calculated parameters in the source domain is illustrated with Figure 6. As described early, nine configurations of parameters represented by values of *c* and gfor the instances in the source domain are trained with instances in the source domain to generate nine SVM classifiers where the instances in the source domain corresponding to the category in the target domain is taken as instances in the source domain, and these nine SVM classifiers are organized in decreasing order of classification performance on the instances of the source domain.

In Figure 6, the color corresponds to SVM classifiers' performance on its respective domain, where red represents the best performances and blue the worst. The three best configurations determined in this order for the source domain are combined with instances in the target domain to generate three SVM classifiers for the target domain. The classifier delivering the best classification performance in the instances from the target domain will be selected as the optimal SVM classifiers for classifying testing images.

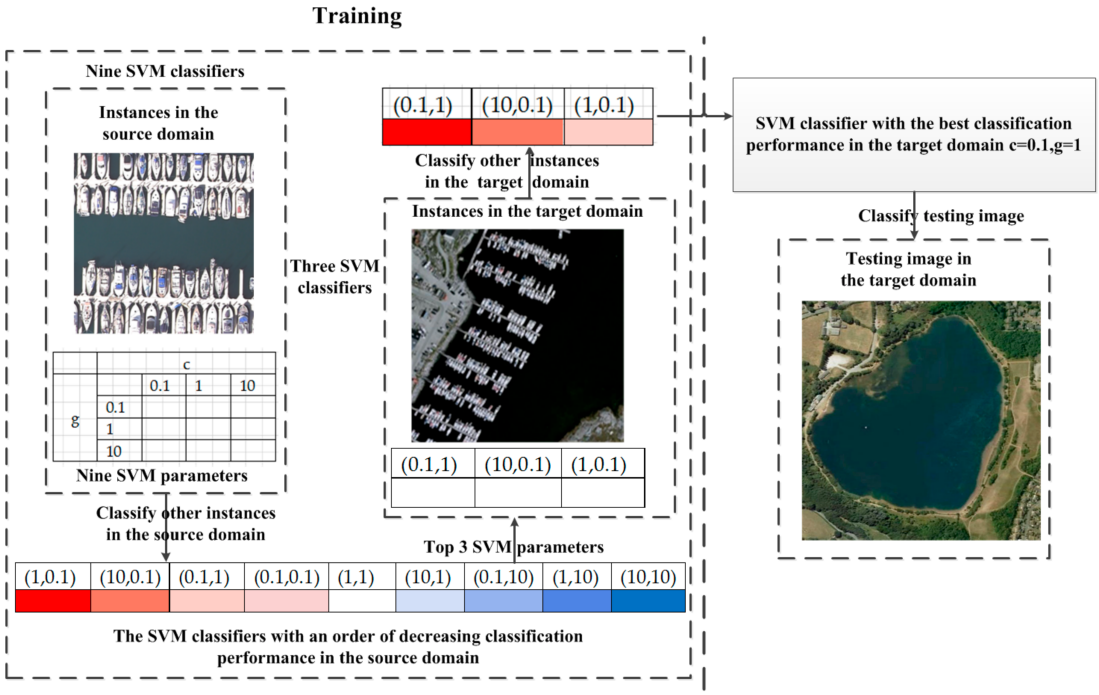


Figure 6. The sketch map of transferring SVM parameters from the source domain to the target domain. The color indicates the performance of the classifier with a decreasing order (from dark red to dark blue).

The methodology can be extended from nine calculated configurations shown in Figure 6, to 25 or more calculated configurations. Corresponding to this situation, more than three SVM classifiers with the best classification performance in the source domain will be applied to the target domain to reduce the search space. The research hypothesis is that the transferred knowledge, in the form of parameters with a good performance within the source domain can bring performance gains with less effort for

learning process in the target domain. Therefore, the process of parameter tuning in the target domain can be facilitated, reducing the distribution bias between instances in the source and target domain.

3. Experiments and Results

3.1. Descriptions of Experimental Data

Two types of publicly-available datasets are used as instances in the source domain, namely AID [62] and the UC MERCED dataset [63]. The UC MERCED dataset was manually extracted from aerial orthoimagery and downloaded from the United States Geological Survey (USGS) National Map. This dataset consists of 21 challenging categories with 100 images per class. The images have a resolution of 30 cm in the RGB color space with a size of 256 × 256. Categories such as freeway, forest, parking lot, and agriculture, and so on, will be used. Some samples of each class are shown in Figure 7.

Figure 7. Sample images of the UC MERCED dataset in the source domain.

AID is a new large-scale aerial image dataset, by collecting instances from Google Earth imagery. The new dataset is made up of 30 aerial scene types and all the images are labelled by the specialists in the field of remote sensing image interpretation. In all, the AID dataset has a number of 10,000 labeled images of 30 classes. Moreover, all the instances per each class in AID are carefully chosen from different countries and regions around the world, mainly in China, the United States, England, France, Italy, Japan, Germany, etc., and they are extracted from different times and seasons with different imaging conditions, which may increase the intra-class diversities of the data. Some samples of each class are shown in Figure 8.

The images of the target domain consist of two different types of images, namely the WHU-RS [64] dataset and the SIRI-WHU [65] dataset. The WHU-RS dataset is a new publicly-available dataset where all the images are collected from Google Earth (Google Inc., Mountain View, CA, USA). It consists of high-resolution satellite scenes of 19 categories. There are 50 images of size 600 × 600 pixels for each class. Sample images of each class in this dataset are shown in Figure 9. The categories in the WHU-RS dataset are all included in the AID dataset.

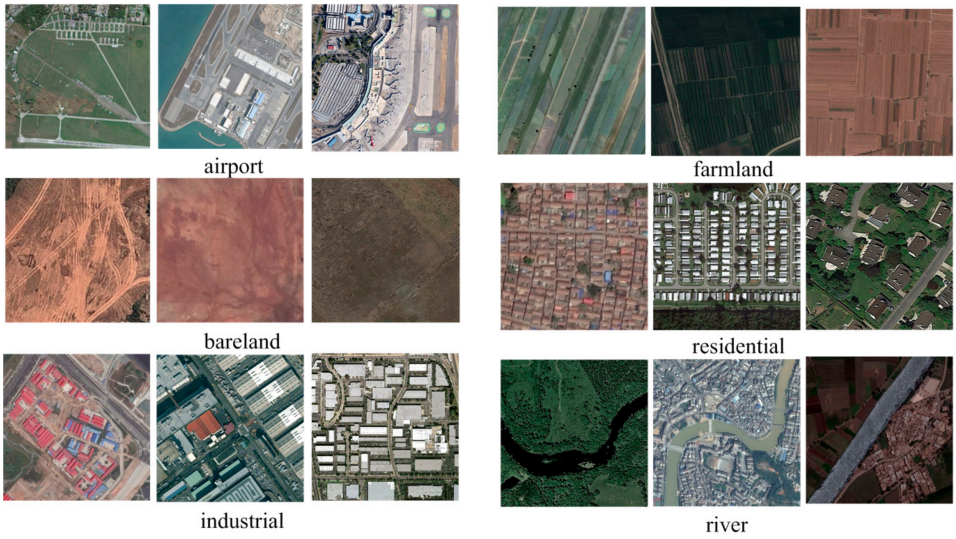


Figure 8. Sample images of the AID dataset used in the source domain.

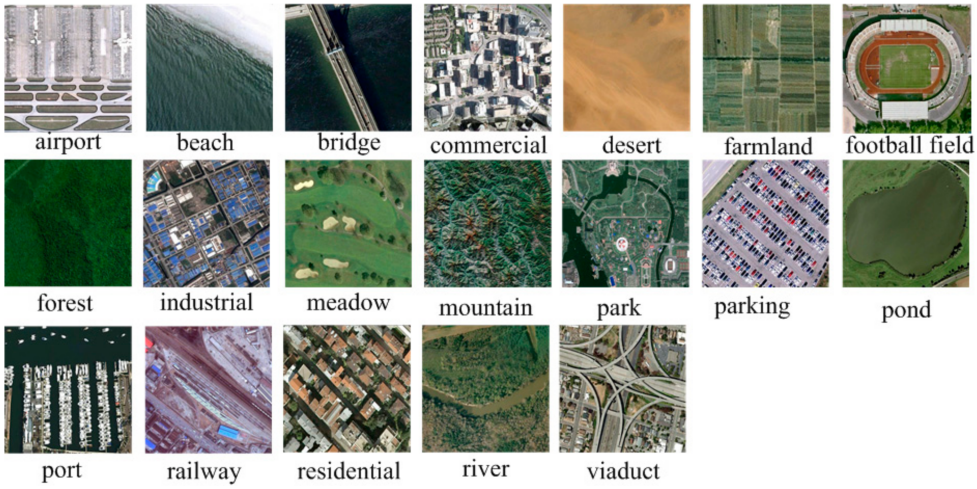


Figure 9. Sample images of the WHU-RS dataset in the target domain with 19 categories.

The SIRI-WHU dataset is acquired from the USGS, covering Montgomery Country, Ohio, and USA. The spatial resolution of this image is 0.6 m. The large image to be annotated is of 10000 × 9000 pixels, as shown in Figure 10a. There are, mainly, six classes for classification: farmland, forest, freeway, parking lot, residential, and other categories mainly consisting of rivers. The original image is converted into 150 × 150 pixel sub images for scene classification, as shown in Figure 10b.

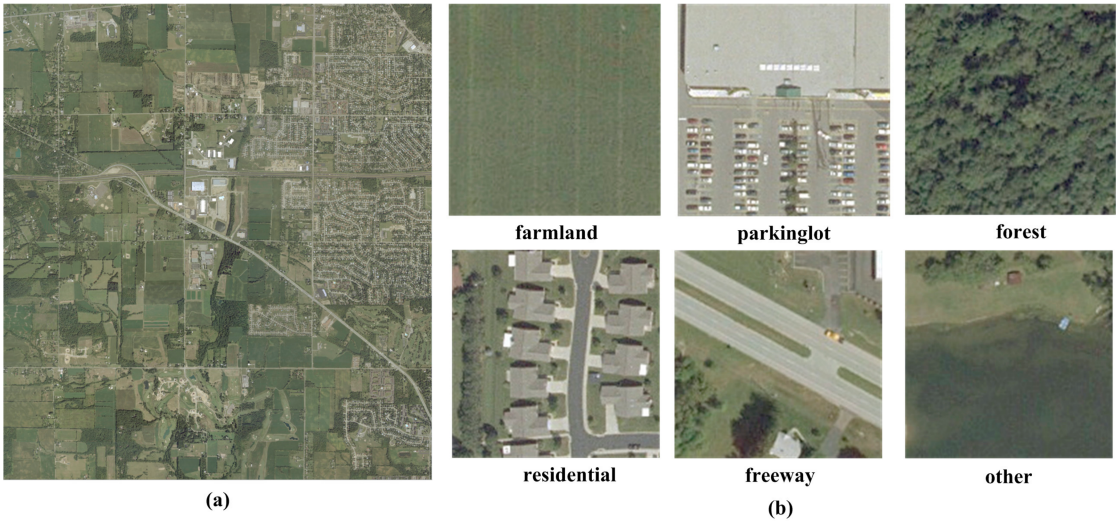


Figure 10. Sample images of the WHU-RS dataset in the target domain with 19 categories. (**a**) Original image; and (**b**) sample images of six land-use categories.

3.2. Experimental Setup

In our experiment, all images are uniformly sampled with a patch size and spacing of eight and four pixels, respectively, to extract DenseSIFT features for generating BOVW representations. To test the stability of the proposed cross-domain scene classification method, the color-boosted saliency-guided rotation invariant BOVW representation with parameter transfers is executed five times by a random selection of instances in the target domain. The instances consist of two parts: all instances in the source domain and a few instances in the target domain. The segmented images of the target domain, except a few utilized instances, will be considered as testing images. Fifty-percent of the instances are used for training and the other 50% for optimal parameter settings. The radial basis function (RBF) kernel has been used in SVM classifiers since it has been proved to deliver better performance in linearly-inseparable instances than other kernels.

According to experience, we will set some parameters in the CBGCSRD method as $\alpha = 0.542$, $\beta = 0.780$, $\gamma = 0.313$, and ${\sigma_s}^2 = 0.4$. Eight parameters, including the multiplied factor k, the parameters in the SVM classifier c cand g, the number of samples in each category No, the number of top configurations t in transferring from the source domain to the target domain, and the codebook size s will be calculated by cross-validation for optimal parameter settings. The calculated optimal parameter settings for the WHU-RS and SIRI-WHU datasets are shown in Table 1.

Table 1. The optimal parameter settings for two target images.

Dataset	Optimal Parameter Settings
WHU-RS	$k = 250, c = 2^{-1}, g = 1, No = 22\%, s = 800, t = 11$
SIRI-WHU	$k = 100, c = 1, g = 2^{-1}, No = 22\%, s = 800, t = 11$

The sensitivity analysis on two datasets will be performed when fixing five other optimal parameters and changing only one parameter as shown in Section 4.

The experimental results are compared with three mid-level methods BOVW, probabilistic latent semantic analysis (PLSA) [66], latent Dirichlet allocation (LDA) [67], and several features of convolutional neural network (CNN) architectures, OverFeat [68], CaffeNet [69], GoogleNet [70], Bag of Convolutional Features (BOCF) [71], and transferred pre-trained VGG-S architectures [72]. In order to evaluate the effectiveness of the three contributions, results only without one of the three contributions and three baseline methods global contrast-based salient region detection method [51],

BOVW with concentric circle-based partition strategy [55], and automatic kernel selection [60] are all obtained. The computer environment is based on a personal computer of Asus in Wuhan, Hubei Province, China with an Intel Core i7-6700 from Intel in Beijing, China with 16GB of RAM without using the GPU for calculating the computational time.

3.3. Results of the WHU-RS Dataset

As can be seen in Table 2, the proposed method outperforms other state-of-the-art methods in terms of classification accuracy with limited instances in the target domain, even some CNN architectures. The CNN architectures, including transferred pre-trained VGG-S and BOCF output classification accuracies close to that of the proposed method because high similarity between instances from the source and target domain, may increase the number and the discriminative ability of the training dataset for CNN architectures. However, all methods based on mid-level representations, such as BOVW, LDA, and PLSA, deliver poor performance because of inadequate descriptions of complex scenes in the WHU-RS dataset.

Table 2. Comparison with state-of-the-art methods in the WHU-RS dataset.

Method	Accuracy (Mean ± Std)
BOVW	72.21 ± 1.55
LDA	70.91 ± 1.37
PLSA	71.25 ± 1.44
OverFeat	94.29 ± 0.69
CaffeNet	95.01 ± 0.71
GoogleNet	94.79 ± 0.74
Transferred pre-trained VGG-S	95.46 ± 0.48
BOCF	94.71 ± 0.55
Color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer	95.82 ± 0.68

As can be seen in Figure 11, the peak classification accuracy of the proposed method outperforms that of all other state-of-the-art methods. It can also be noted in Figure 11 and Table 2 that the proposed method is more stable in classification accuracy than most state-of-the-art methods except transferred pre-trained VGG-S model and BOCF since it is difficult to detect salient regions in some scenes of similar texture with the proposed CBGCSRD method.

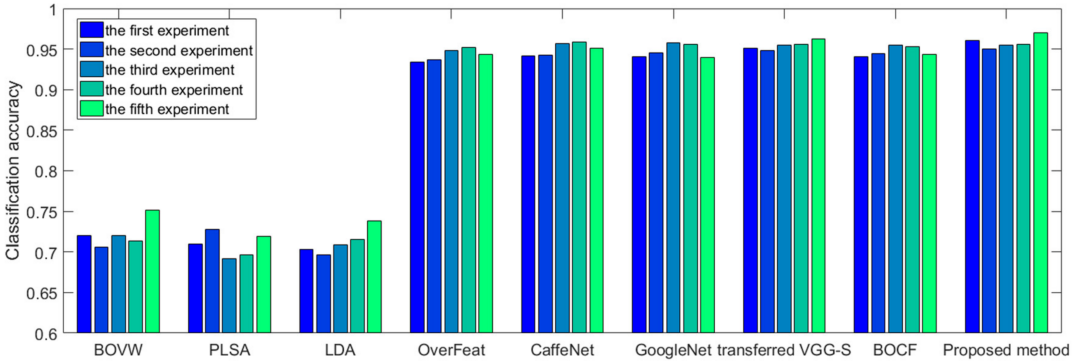


Figure 11. The classification accuracy of five experiments on the proposed method and the state-of-the-art methods in the WHU-RS dataset.

As can be shown in Table 3, the CBGCSRD method, rotation invariant BOVW representation and the transfer of SVM parameters from the source to the target domain are all effective in improving the classification accuracy, while the classification accuracy of the rotation invariant BOVW representation improves the most since it can decrease the effect of the rotation transformations in instances.

The classification accuracy of methods without three contributions is about 8% lower than that of the proposed method, demonstrating the effectiveness of the proposed cross-domain scene classification method since it solves the three problems in Section 1. It can also be seen from Table 3 that the color-boosted method plays an important role in the salient region detection since the salient region is more accurate because of higher contrast between different regions of images. Table 3 also shows that the proposed rotation invariant BOVW representation is more effective in improving the classification accuracy compared with BOVW representation with concentric circle-based partition strategy since there exists a loss in the spatial layout information on the BOVW representation with the concentric circle-based partition strategy compared with that with SPM.

Table 3. Comparison with methods without three contributions in the WHU-RS dataset.

Method	Accuracy (Mean ± Std)
Method without three contributions	87.75 ± 0.95
Only without the CBGCSRD method	92.79 ± 0.84
Only without the color-boosted method, but with global contrast-based salient region detection	91.08 ± 0.93
Only without rotation invariant BOVW representation	92.06 ± 0.88
Only without rotation invariant BOVW representation, but with a concentric circle-based partition strategy	90.31 ± 0.97
Only without transferring SVM parameters from the source to the target domain	93.61 ± 0.96
Only without transferring SVM parameters, but with automatic kernel selection	93.88 ± 0.87
Color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer	95.82 ± 0.77

As can be seen in Table 4, the proposed color boosted saliency-guided rotation invariant BOVW representation with parameter transfers method outperforms other two methods in terms of computational time since it reduces the effort of searching for the optimal parameters or kernels from a larger number of candidate parameters or kernels for the target domain. It can also be seen from Tables 3 and 4 that although the automatic kernel selection increases a slight classification accuracy, it increases the computational time by 3 h compared with methods without transferring SVM parameters.

Table 4. The superiority of the color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer method in reducing the computational time in the WHU-RS dataset compared with state-of-the-art methods.

Method	Computational Time (h)
Only without transferring SVM parameters from the source to the target domain	25.37
Only without transferring SVM parameters, but with automatic kernel selection	28.45
Color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer	13.50

The classification accuracy in each category and confusion matrix of the color-boosted saliency-guided rotation invariant BOVW representation with the parameter transfer method or methods without only one of three contributions are shown in Figures 12 and 13, respectively. As can be seen in Figure 12, the proposed method outperforms methods without one of three contributions in all categories, except for desert, farmland, forest, meadow, and mountain categories, since scenes of these categories demonstrate low contrast between different regions, leading to inaccurate salient region detection results and classification performance. The CBGCSRD method, rotation invariant BOVW representation, and the transfer of SVM parameters from the source and the target domains

are all effective in improving the classification accuracy of all categories to different degrees while the classification accuracy of rotation invariant BOVW representation improves the most in all categories due to reduced effect of the rotation transformation on the classification.

As can be seen in Figure 13a, the proposed method achieves good classification performance in almost all categories, except desert, farmland, forest, meadow, and mountain since categories, with higher spectral contrast are easier for salient region detection, which may better reduce the effect of the background on the BOVW representation. As can also be seen in Figure 13b, the classification accuracies of all categories, except the five categories mentioned above, decrease because BOVW has been performed on the salient region rather than the whole image, reducing the effect of backgrounds. We can also see in Figure 13c,d that the rotation invariant BOVW representation and transfer onto SVM parameters improve classification in all categories, but transfer of the SVM parameters improves the classification accuracy less because of the similarity between instances of the WHU-RS dataset and those in the AID dataset.

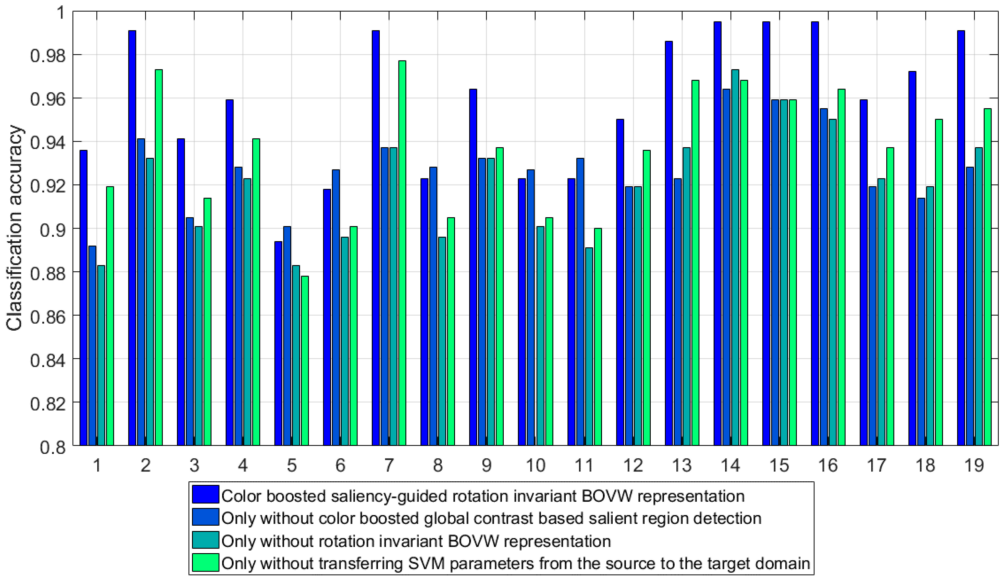
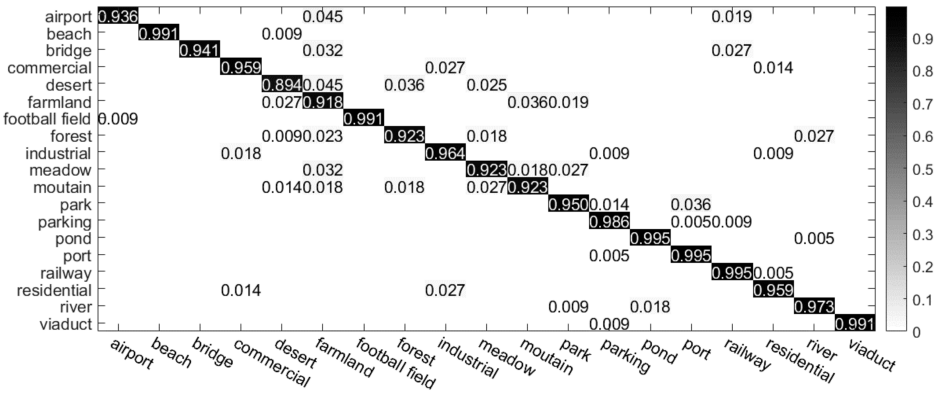


Figure 12. Producers' accuracies with the WHU-RS dataset for four different methods in Table 3. The class labels are assigned as follows: 1 = airport, 2 = beach, 3 = bridge, 4 = commercial, 5 = desert, 6 = farmland, 7 = football field, 8 = forest, 9 = industrial, 10 = meadow, 11 = mountain, 12 = park, 13 = parking, 14 = pond, 15 = port, 16= railway station, 17 = residential, 18 = river, and 19 = viaduct.



(**a**)

Figure 13. *Cont.*

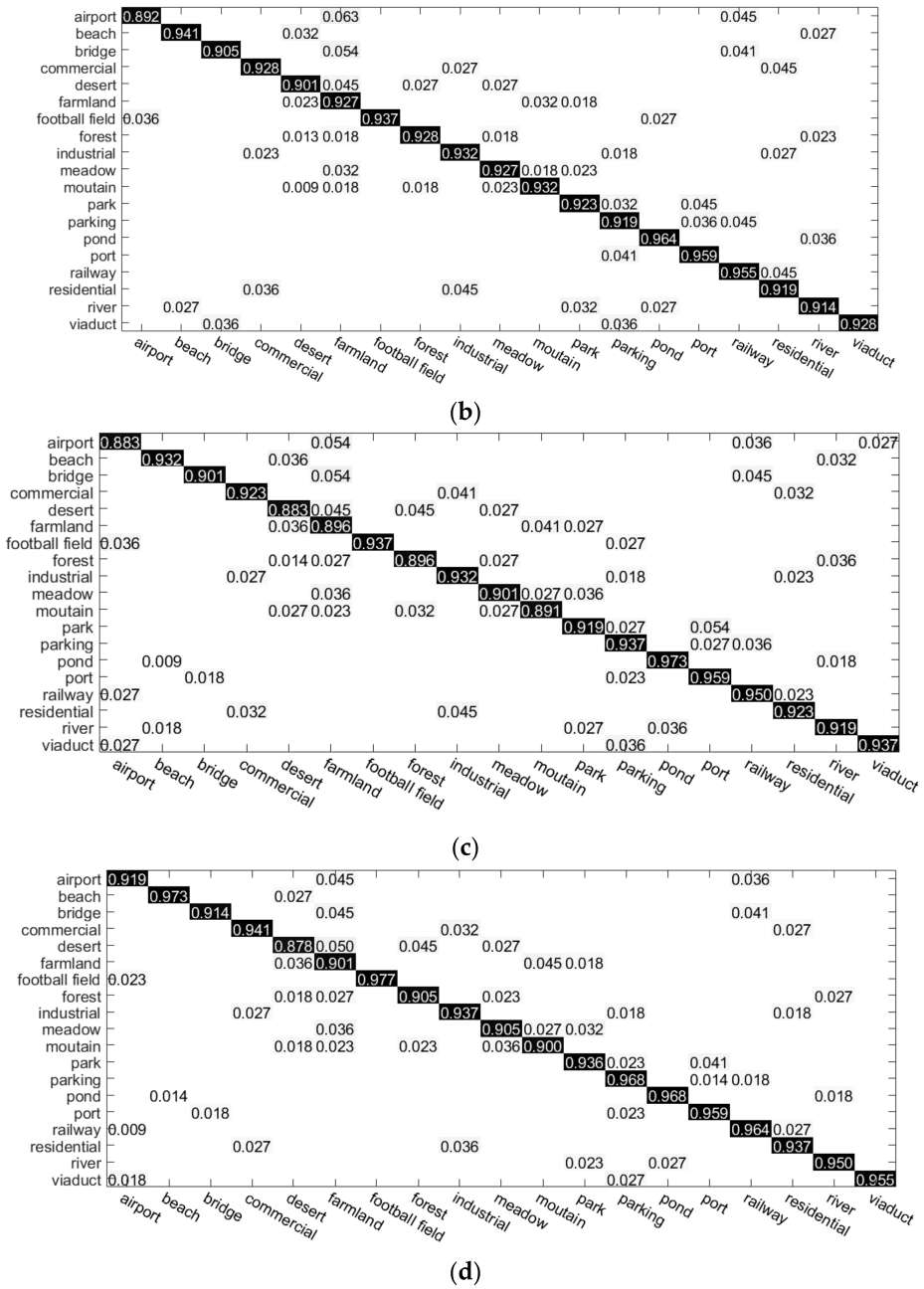


Figure 13. The confusion matrix for different methods in Table 3. (**a**) Color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer method; (**b**) only without CBGCSRD method (**c**) only without rotation invariant BOVW representation; and (**d**) only without transferring SVM parameters from the source to the target domain.

3.4. Results of the SIRI-WHU Dataset

The color-boosted saliency-guided rotation invariant BOVW representation with parameter transfers outperforms other state-of-the-art methods in terms of classification accuracy as shown in Table 5 since it solves the problems mentioned above. Three CNN architectures, namely OverFeat, CaffeNet, and GoogleNet, deliver lower classification accuracy than the proposed method. The transferred pre-trained VGG-S and BOCF output close accuracies to the proposed method because they are more discriminative in classifying categories with high intra-class variability. As can also be

seen in comparison between Tables 2 and 5, the CNN architectures demonstrate lower accuracy in Table 5 since the instances in the source domain and those in the SIRI-WHU dataset are highly different.

Table 5. Comparison with state-of-the-art methods in the SIRI-WHU dataset in classification accuracy.

Method	Accuracy (Mean ± Std)
BOVW	68.80 ± 1.12
LDA	69.91 ± 1.24
PLSA	68.25 ± 1.34
OverFeat	90.29 ± 0.75
CaffeNet	90.31 ± 0.62
GoogleNet	89.99 ± 0.72
Transferred pre-trained VGG-S	91.12 ± 0.53
BOCF	90.70 ± 0.58
Color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer	91.65 ± 0.63

As can be seen in Figure 14, the maximum classification accuracy of the proposed method is the highest in all nine methods. With different target instances, the proposed method delivers relatively stable classification performance comparable to some CNN architectures as shown in Table 5 and Figure 14. However, when faced with scenes with low spectral contrast between different regions, the CBGCSRD method may deliver poor performance, which may have a negative influence on the stability of the proposed method.

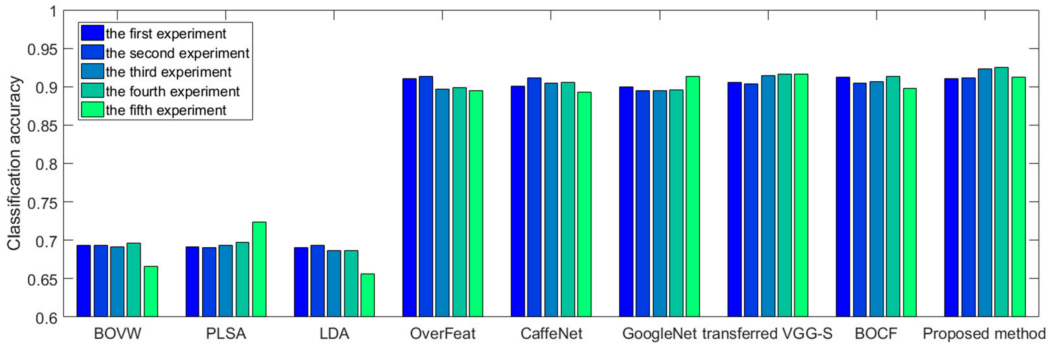


Figure 14. The classification accuracy of five experiments on the proposed method and the state-of-the-art methods in the SIRI-WHU dataset.

As can be seen in Table 6, three contributions including CBGCSRD, rotation invariant BOVW representation, and transfer of SVM parameters from the source to the target domain are all proved to be effective in improving the classification accuracy since these three strategies constrain the problem of background information, rotation transformation, and feature difference between scenes of the source and target domain. However, the transfer onto SVM parameters plays the most important role in increasing the classification accuracy since it finds the optimal SVM parameters from the source domain for the target domain. Methods without three contributions deliver 7% lower classification accuracy than the proposed method, demonstrating the effectiveness of three contributions to cross-domain scene classification since it handles three problems mentioned above.

As can be seen in Table 7, the proposed method also reduces about 8 to 10 h than other two methods because of reduced search space for optimal parameter settings. It can also be noted that the automatic kernel selection method can increase classification accuracy with loss of some computational time.

Table 6. Comparison with methods concerning three contributions in the SIRI-WHU dataset.

Method	Accuracy (Mean ± Std)
Method without three contributions	84.64 ± 0.80
Only without the CBGCSRD method	89.68 ± 0.73
Only without the color-boosted method, but with global contrast based salient region detection	88.25 ± 0.85
Only without rotation invariant BOVW representation	88.86 ± 0.74
Only without rotation invariant BOVW representation, but with a concentric circle-based partition strategy	87.18 ± 0.83
Only without transferring SVM parameters from the source to the target domain	87.77 ± 0.85
Only without transferring SVM parameters, but with automatic kernel selection	88.25 ± 0.71
Color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer	91.65 ± 0.66

Table 7. The superiority of the color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer method in reducing computational time in the SIRI-WHU dataset compared with other state-of-the-art methods.

Method	Computational Time (h)
Only without transferring SVM parameters from the source to the target domain	16.76 ± 0.87
Only without transferring SVM parameters, but with automatic kernel selection	18.82 ± 0.82
Color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer	8.92 ± 0.77

The classification accuracy in each category and the confusion matrix of the proposed method or methods without three contributions are shown in Figures 15 and 16, respectively. As can be seen in Figure 15, the proposed method outperforms that without three contributions in all categories except farmland and forest since these scenes with low contrast between different regions do not perform well in the CBGCSRD method. The rotation invariant BOVW representation, and the transfer of SVM parameters from the source domain to the target domain can increase the classification accuracy of all categories to varying degree while the transfer of SVM parameters increases the highest classification accuracy in all categories which are highly different from those in the source domain. The CBGCSRD method increases the classification accuracy of categories except farmland and forest categories. It can also be noted in Table 5 that the proposed rotation invariant BOVW representation and color-boosted method demonstrate higher superiority in increasing the classification accuracy for images of the target domain since they solve the problems of the BOVW representation with a concentric circle-based partition strategy and global contrast-based salient region detection similar to Section 3.3.

As can be seen in Figure 16, the proposed method delivers performance better than 0.9 in freeway, other, parking lot and residential categories since these categories are with higher color contrast between different regions. The major confusion usually occurs to the scenes farmland and forest, forest and freeway, farmland and other, forest and residential, and so on. The CBGCSRD method increases the accuracy in all categories, except categories delivering poor classification performance in the salient region detection method. The rotation invariant BOVW representation can help to enhance the performance in categories with rotation transformations while the transfer of SVM parameters can improve the classification performance of categories different from those in the source domain.

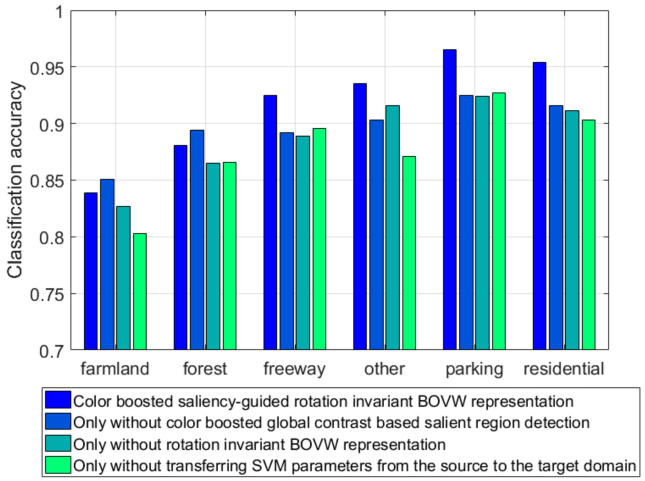


Figure 15. Producers' accuracies with the SIRI-WHU dataset for four different methods.

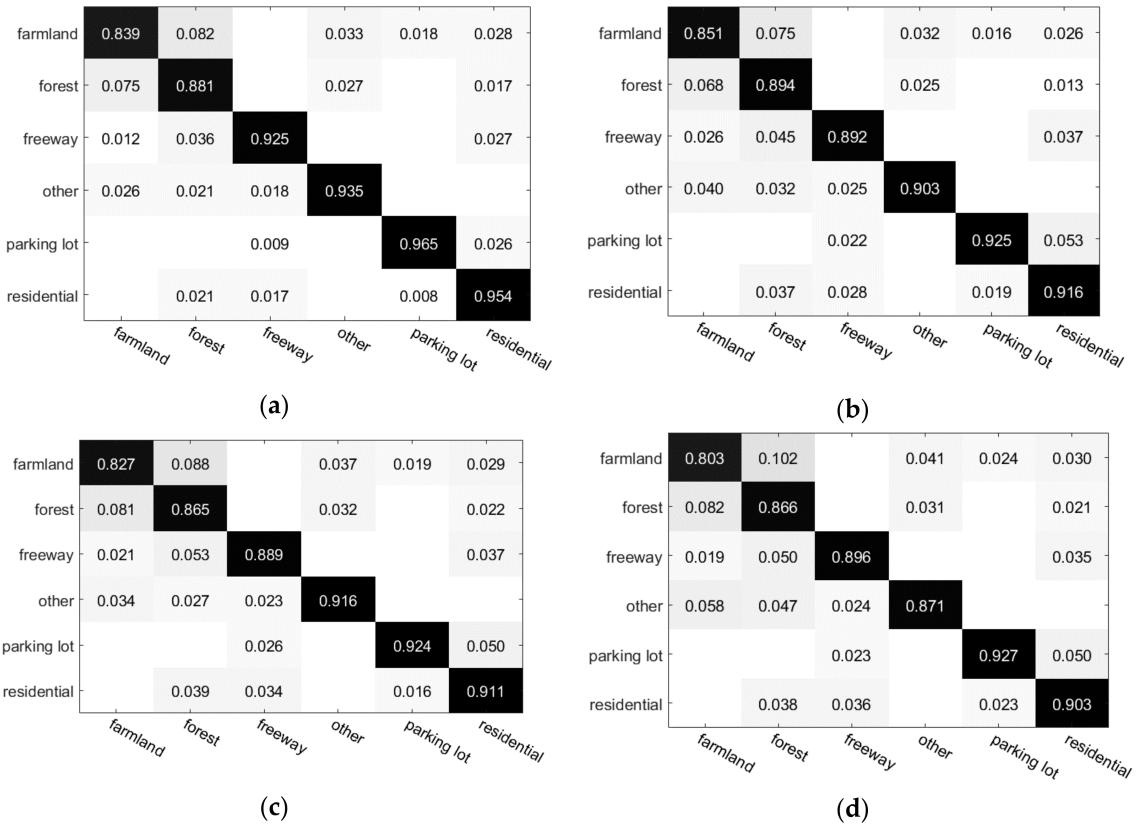


Figure 16. Confusion matrix showing the classification performance with the SIRI-WHU dataset: (**a**) color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer; (**b**) only without the CBGCSRD method; (**c**) only without rotation invariant BOVW representation; and (**d**) only without transferring SVM parameters from the source to the target domain.

The annotated map for the SIRI-WHU image is also obtained as shown in Figure 17b. As can be seen in Figure 17, the major confusion may occur to the border of different categories since these images may demonstrate a mixture of objects of different categories. When faced with scenes with a mixture of objects of different categories, the scenes are usually assigned to the labels covering the largest area in the image, which may increase the difficulty of salient region detection.

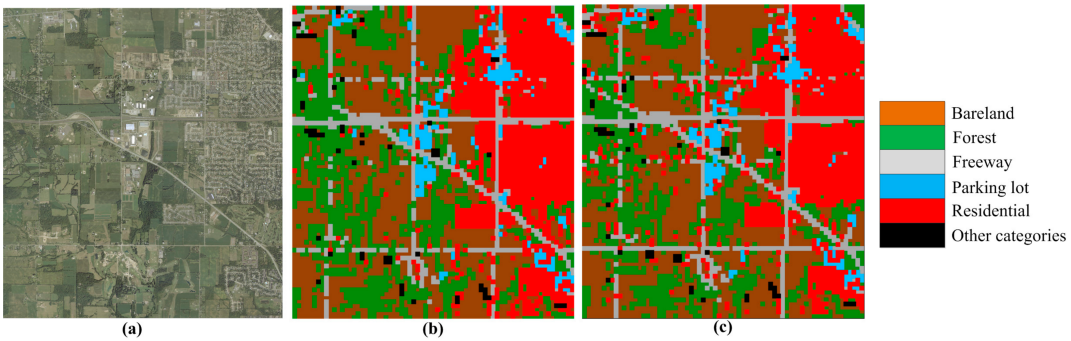


Figure 17. Semantic annotation of SIRI-WHU dataset. (**a**) Original image; (**b**) annotation result of SIRI-WHU dataset with color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer corresponding to Table 5; and (**c**) ground truth image.

4. Discussion

4.1. Parameter Sensitivity Analysis

Four parameter settings, multiplied factor k, codebook size s, the percentage of instances in the target domain No, and several top configurations in the source domain t will all have an effect on the cross-domain classification accuracy. In order to evaluate the effectiveness of color-boosted saliency-guided rotation invariant BOVW representation with parameter transfers method for cross-domain scene classification, we analyze the sensitivity of these four parameter settings for two datasets mentioned in Section 3.1.

4.1.1. Influence of Multiplied Factor k in the Color-Boosted Salient Global Contrast Based Region Detection Method

In order to investigate the sensitivity of the proposed method in relation to the multiplied factor k, other parameters will be fixed as optimal parameter settings in each dataset as shown in Table 1. The multiplied factor k was then varied from the range of [5, 10, 25, 50, 100, 250, and 500] for two datasets.

As shown in Figure 18, the classification accuracy increases at first before decreasing gradually. The peak of classification accuracy reflects the most suitable multiplied factor k. The proposed method is more sensitive to the multiplied factor k in the WHU-RS dataset compared with the SIRI-WHU dataset.

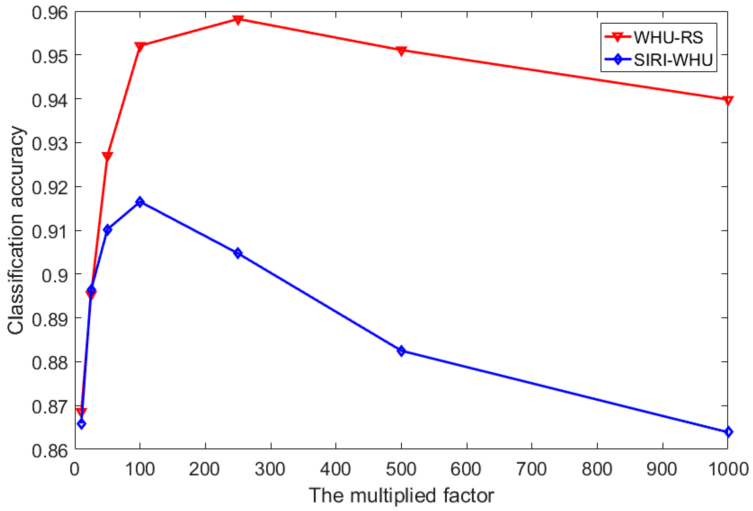


Figure 18. Effect of the multiplied factor k on the classification accuracy in two datasets.

4.1.2. Influence of the Codebook Size *s*

Different sizes of visual vocabularies were tested on different sizes from 50 to 950 at intervals of 150 for both datasets since the number of visual words has an effect on classification accuracy. To study the sensitivity of the proposed method in codebook size *s*, other parameters are kept constant as shown in Section 4.1.1.

From Figure 19, it is notable that the trends for two datasets are similar. With the increase in codebook size *s*, the classification accuracy improves gradually. However, when the codebook size *s* is too large, the classification accuracy may improve only slightly. The WHU-RS dataset is more sensitive to the codebook size *s* than the SIRI-WHU dataset.

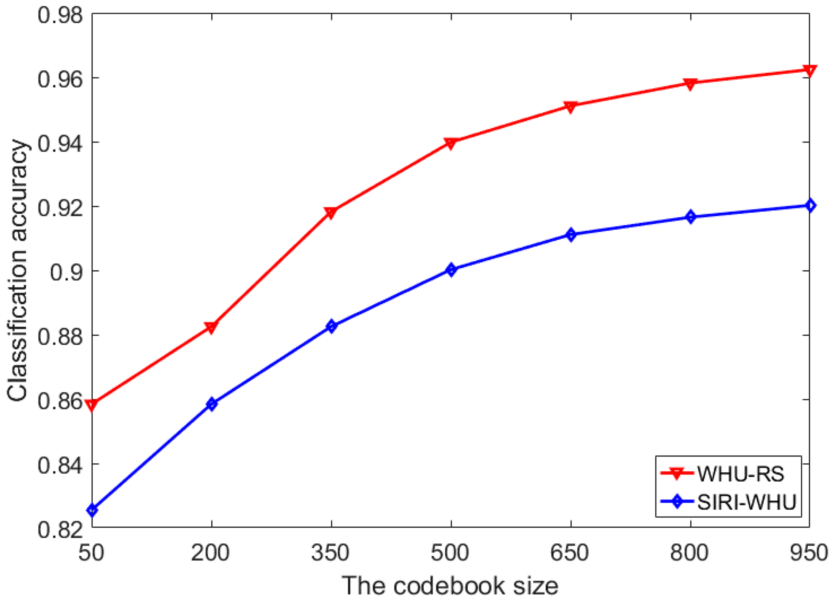


Figure 19. Effect of the codebook size *s* on the classification accuracy in two datasets.

4.1.3. Influence of the Percentage of Instances *No* in the Target Domain on Methods with or Without Parameter Transfer

The classification accuracy of methods with or without parameter transfer will be limited by the percentage of instances in the target domain *No*. In order to test the sensitivity of the proposed method and the method without parameter transfer in relation to the percentage of instances in the target domain *No*, the percentage of instances in the target domain is varied from [0.02, 0.06, 0.10, 0.14, 0.18, 0.22, and 0.26]. The trends of classification accuracy with the percentage of instances from the target domain in two datasets for different methods are similar, as displayed in Figure 20. The classification accuracy improves gradually before improving little with the increased percentage of instances in the target domain *No*. However, the methods without parameter transfer are more sensitive to the percentage of instances in the target domain *No* than those with parameter transfer since an inadequate number of instances in the target domain may lead to indiscriminative classifiers but parameter transfer may make for this deficiency to some degree. When the percentage *No* is 0.26, the classification accuracies of methods with or without parameter transfer are close since the number of instances are enough for training a discriminative classifier.

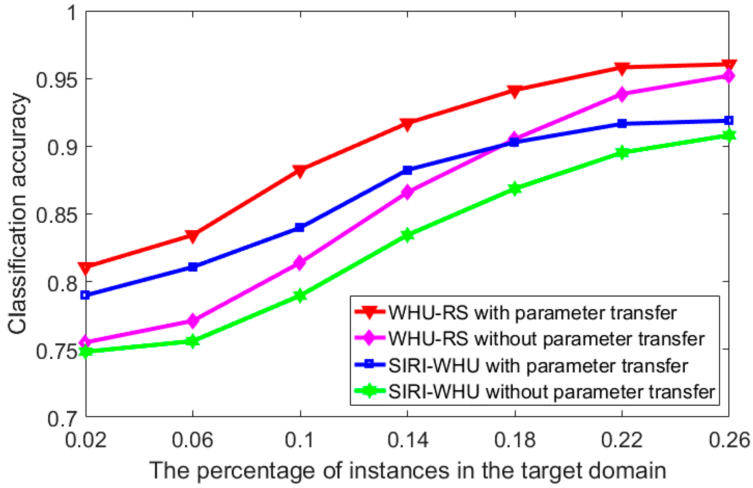


Figure 20. Effect of the percentage of instances in the target domain on classification accuracy in two datasets.

4.1.4. Influence of the Number of Configurations Used for the Target Domain *t*

In order to study the effect of the number of configurations used for the target domain *t* for the proposed method, the number of configurations *t* is varied from the range of [1, 3, 5, 7, 9, 11, and 13] in the two datasets. The overall trends of classification accuracy in the two datasets are similar with the increase in the number of configurations *t* as shown in Figure 21. The classification accuracy increases when the number of configurations *t* is below 11, then it improves only slightly.

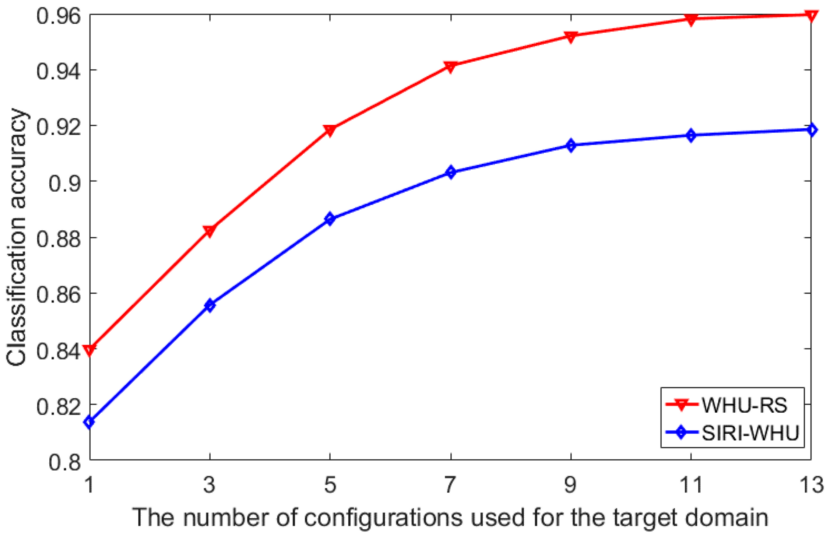


Figure 21. Effect of the number of configurations used for the target domain *t* on the classification accuracy in two datasets.

4.2. Analysis of Results in Two Target Images

Extensive experiments show that the proposed method which integrates CBGCSRD method, rotation invariant BOVW representation with transferring SVM parameters from the source to the target domain is very effective in the cross-domain scene classification method.

The existing BOVW representation is calculated from the whole image, which may be affected by the background information. The color-boosted saliency-guided rotation invariant BOVW

representation with parameter transfer concentrates on the detected salient regions rather than the whole image. Experimental results of the WHU-RS and SIRI-WHU datasets indicate that the color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer method is competitive with the state-of-the-art methods in terms of classification accuracy, even some CNN architectures, as displayed in Figures 11–17. However, the proposed method delivers poor performance in some example images of the WHU-RS dataset shown in Figure 22.

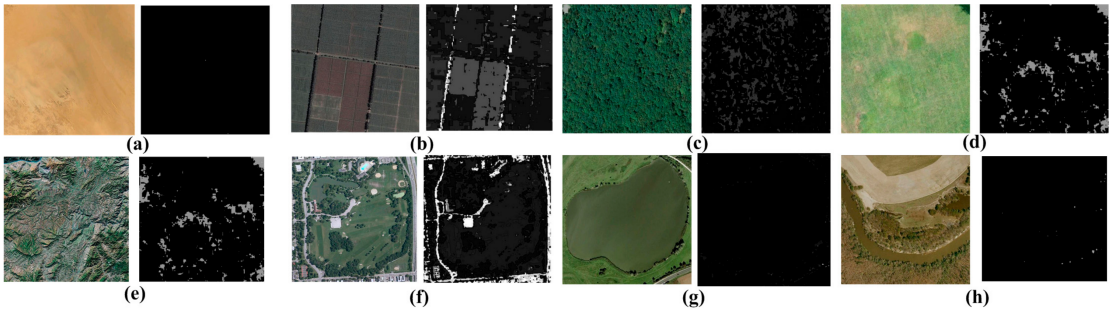


Figure 22. Several examples and their corresponding color boosted saliency maps of different categories misclassified by the color boosted saliency-guided rotation invariant BOVW representation with parameter transfer in the WHU-RS dataset: (**a**) desert, (**b**) farmland, (**c**) forest, (**d**) meadow, (**e**) river, (**f**) park, (**g**) pond, and (**h**) mountain.

As shown in Figure 22a–h display some misclassified example images of different categories. These images including eight different categories demonstrate similar color in different regions of the scene, leading to low saliency in almost the whole image. Therefore, there exist nearly no salient regions in these scenes, which may lead to insufficient descriptors for descriptions of images.

As can also be seen in the Figures 12 and 13, the rotation invariant BOVW representation can increase the classification accuracy of categories significantly affected by the rotation transformation such as residential, parking, port, viaduct, and so on while the classification accuracies of categories including forest, desert, and meadow less affected by rotation transformation change little. Moreover, as the scenes of the WHU-RS dataset are similar to those in the AID dataset, the transfer of SVM parameters increases a little accuracy in all categories.

Similarly, as can be seen in Figure 23a–g, images with low color contrast between different regions may deliver poor salient region detection results, which may result in poor classification performance.

As we can see in Figures 14–17, the rotation invariant BOVW representation can also increase the classification accuracy of categories with a reduced effect of rotation transformation on image representation. However, the transfer of SVM parameters from the source to the target domain plays the most important role in improving the classification accuracy since the scenes of some categories such as other, residential, and parking lots in the SIRI-WHU dataset are highly dissimilar to those in the source domain. The transfer of SVM parameters from the source to the target domain can help to increase the performance of cross-domain scene classification.

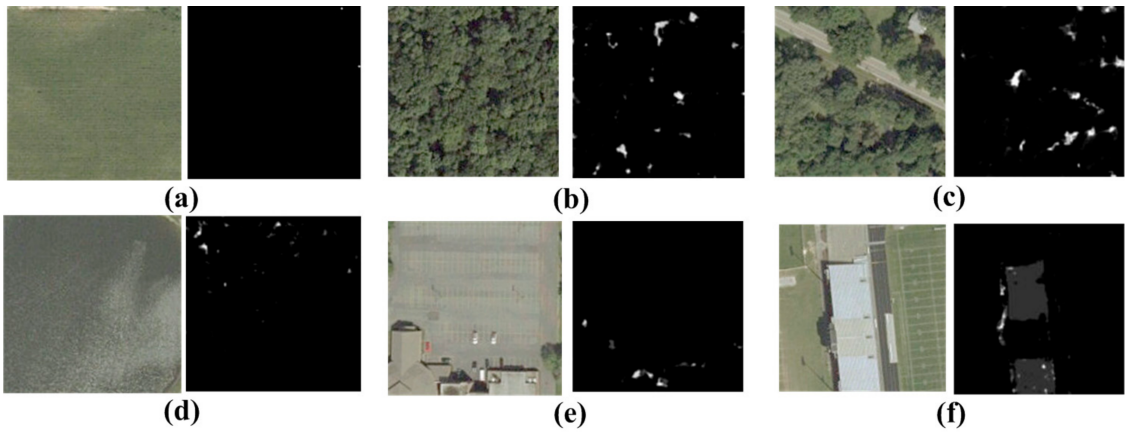


Figure 23. Several examples of different categories misclassified by the color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer in the SIRI-WHU dataset: (**a**) farmland, (**b**) forest, (**c**) freeway, (**d**) other, (**e**) parking lot, and (**g**) residential.

4.3. Strengths and Limitations

A cross-domain scene classification method based on the CBGCSRD method, rotation invariant BOVW representation, and transfer of SVM parameters from the source domain to the target domain is proposed in this study. This method has been successfully applied to two target datasets with limited instances in the target domain. The main advantage of the proposed approach is increased classification accuracy in the cross-domain scene classification method and transferring SVM parameters from the source domain and the target domain. Experimental results show that this method can achieve an overall classification accuracy of 95.82% and 91.65% in different target datasets with limited instances in the target domain and reduced computational time and outperforms most state-of-the-art scene classification methods, and even some CNN architectures.

In conclusion, the proposed method can achieve excellent classification accuracy in categories with relatively high color contrast since the effect of background information has been reduced. The proposed method is also robust to rotation transformation due to decreased influence of rotation transformation on rotation invariant BOVW representation. The method can reduce the computational cost and manual intervention in labeling target instances by transferring parameters from the source domain to the target domain.

However, it may deliver poor classification performance in categories with low contrast since these categories demonstrate similar color in different regions of the scene, leading to low saliency in almost the whole image by the color-boosted salient region detection method. Therefore, there exist nearly no salient regions in these scenes, which may lead to insufficient SIFT descriptors for descriptions of images.

5. Conclusions

In this paper, a color-boosted saliency-guided rotation invariant BOVW representation with parameter transfer is proposed to cross-domain scene classification in order to solve existing problems, namely the influence of background on salient region detection and rotation transformation on BOVW representation, and transfer onto SVM parameters from the source to the target domain to reduce the number of required instances in the target domain. In order to solve problems mentioned above, CBGCSRD method is proposed to detect salient regions for descriptor extraction with higher color contrast between different segmented regions. To better reduce the effect of rotation transformation on the classification accuracy, the existing BOVW representation has been modified by sorting the representations of each patch of images divided into eight patches with the same area. The several configurations performing the best in the source domain will be applied to the target domain to

complete the transfer of SVM parameters from the source domain to the target domain to reduce the required instances in the target domain and computational cost. The experiments on two different datasets as the target domain show the following conclusions:

(1) The CBGCSRD method, rotation invariant BOVW representation, and transfer of SVM parameters from the source to the target domain are all effective in improving the classification performance while the rotation invariant BOVW representation and transfer of SVM parameters play the most important roles in the WHU-RS and SIRI-WHU dataset, respectively, since the WHU-RS dataset is similar to the AID dataset, but the SIRI-WHU dataset is dissimilar to instances in the source domain to some degree. The parameter transfer not only improves classification accuracy, but also decreases computational time.

(2) The color boosted saliency-guided rotation invariant BOVW representation with parameter transfer method outperforms most previous methods in terms of classification accuracy with limited instances in the target domain, even some well-known CNN architectures, demonstrating the superiority in the proposed method in terms of classification accuracy.

(3) The CBGCSRD method and proposed rotation invariant BOVW representation demonstrate superiority over existing global contrast based salient region detection and BOVW with a concentric circle-based partition strategy in improving the classification accuracy. The SVM parameter transfer not only improves the classification accuracy, but also decreases the computational time with limited instances in the target domain.

In our future work, in order to suppress the drawbacks, deep CNN features can be used to replace DenseSIFT or other features with no need to detect salient regions in order to generate more discriminative BOVW features for those categories with low contrast. However, the way to combine BOVW with DCNN features need to be explored.

Acknowledgments: The authors would like to thank the editors and the anonymous reviewers for their comments and suggestions. This research was supported by The National Key Research and Development Program of China under grant no. 2016YFB0501403 and The Key Research and Development Program of Jiangxi Province under grant no. 20171BBE50062. The authors would like to thank the USGS for providing the UC_MERCED dataset and the State Key Laboratory in Wuhan University for providing the WHU-RS dataset, SIRI-WHU dataset, and AID dataset. The author would like to thank Professor Cheng from Northwestern Polytechnical University and Doctor Hu from the Wuhan University for providing the source code or implementation details.

Author Contributions: Ruixi Zhu conceived and designed the experiments; Ruixi Zhu and Nan Mo performed the experiments; Li Yan and Ruixi Zhu and Yi Liu analyzed the data; Ruixi Zhu and Nan Mo contributed the use of analysis tools; Ruixi Zhu, Li Yan, and Nan Mo wrote the paper; and Yi Liu helped to prepare the manuscript. All authors read and approved the final manuscript.

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

References

1. Prasad, S.; Bruce, L.M. Decision Fusion with Confidence-Based Weight Assignment for Hyperspectral Target Recognition. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1448–1456. [CrossRef]
2. Bruzzone, L.; Carlin, L. A Multilevel Context-Based System for Classification of Very High Spatial Resolution Images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2587–2600. [CrossRef]
3. Rizvi, I.A.; Mohan, B.K. Object-Based Image Analysis of High-Resolution Satellite Images Using Modified Cloud Basis Function Neural Network and Probabilistic Relaxation Labeling Process. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4815–4820. [CrossRef]
4. Bellens, R.; Gautama, S.; Martinez-Fonte, L.; Philips, W.; Chan, C.W.; Canters, F. Improved Classification of VHR Images of Urban Areas Using Directional Morphological Profiles. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2803–2813. [CrossRef]
5. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]

6. Bratasanu, D.; Nedelcu, I.; Datcu, M. Bridging the Semantic Gap for Satellite Image Annotation and Automatic Mapping Applications. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 193–204. [CrossRef]
7. Bosch, A.; Muñoz, X.; Martí, R. Which is the best way to organize/classify images by content? *Image Vis. Comput.* **2007**, *25*, 778–791. [CrossRef]
8. Li, C.; Wang, J.; Wang, L.; Hu, L.; Peng, G. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery. *Remote Sens.* **2014**, *6*, 964–983. [CrossRef]
9. Weng, Q. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sens. Environ.* **2012**, *117*, 34–49. [CrossRef]
10. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain Adaptation via Transfer Component Analysis. *IEEE Trans. Neural Netw.* **2011**, *22*, 199–210. [CrossRef] [PubMed]
11. Jiang, J.; Zhai, C.X. Instance Weighting for Domain Adaptation in NLP. In Proceedings of the Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 25–27 June 2007; pp. 264–271.
12. Liu, Y.; Li, X. Domain adaptation for land use classification: A spatio-temporal knowledge reusing method. *ISPRS J. Photogramm. Remote Sens.* **2014**, *98*, 133–144. [CrossRef]
13. Tuia, D.; Persello, C.; Bruzzone, L. Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 41–57. [CrossRef]
14. Izquierdo-Verdiguier, E.; Laparra, V.; Gómez-Chova, L.; Camps-Valls, G. Encoding Invariances in Remote Sensing Image Classification With SVM. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 981–985. [CrossRef]
15. Persello, C.; Bruzzone, L. Kernel-Based Domain-Invariant Feature Selection in Hyperspectral Images for Transfer Learning. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2615–2626. [CrossRef]
16. Bruzzone, L.; Persello, C. A Novel Approach to the Selection of Spatially Invariant Features for the Classification of Hyperspectral Images with Improved Generalization Capability. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3180–3191. [CrossRef]
17. Li, X.; Zhang, L.; Du, B.; Zhang, L.; Shi, Q. Iterative Reweighting Heterogeneous Transfer Learning Framework for Supervised Remote Sensing Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2022–2035. [CrossRef]
18. Elshamli, A.; Taylor, G.W.; Berg, A.; Areibi, S. Domain Adaptation Using Representation Learning for the Classification of Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4198–4209. [CrossRef]
19. Tuia, D.; Campsvalls, G. Kernel Manifold Alignment for Domain Adaptation. *PLoS ONE* **2016**, *11*, 1–25. [CrossRef] [PubMed]
20. Yang, H.L.; Crawford, M.M. Spectral and Spatial Proximity-Based Manifold Alignment for Multitemporal Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 51–64. [CrossRef]
21. Yang, H.L.; Crawford, M.M. Domain Adaptation with Preservation of Manifold Geometry for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 543–555. [CrossRef]
22. Sun, H.; Liu, S.; Zhou, S.; Zou, H. Unsupervised Cross-View Semantic Transfer for Remote Sensing Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 13–17. [CrossRef]
23. Sun, H.; Liu, S.; Zhou, S.; Zou, H. Transfer Sparse Subspace Analysis for Unsupervised Cross-View Scene Model Adaptation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2901–2909. [CrossRef]
24. Ye, M.; Qian, Y.; Zhou, J.; Tang, Y.Y. Dictionary Learning-Based Feature-Level Domain Adaptation for Cross-Scene Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1544–1562. [CrossRef]
25. Chaib, S.; Liu, H.; Gu, Y.; Yao, H. Deep Feature Fusion for VHR Remote Sensing Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4775–4784. [CrossRef]
26. Bahirat, K.; Bovolo, F.; Bruzzone, L.; Chaudhuri, S. A Novel Domain Adaptation Bayesian Classifier for Updating Land-Cover Maps with Class Differences in Source and Target Domains. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2810–2826. [CrossRef]
27. Rajan, S.; Ghosh, J. Exploiting Class Hierarchies for Knowledge Transfer in Hyperspectral Data. *Geosci. Remote Sens. IEEE Trans.* **2005**, *44*, 3408–3417. [CrossRef]
28. Chi, M.; Bruzzone, L. Semisupervised Classification of Hyperspectral Images by SVMs Optimized in the Primal. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1870–1880. [CrossRef]

29. Sun, Z.; Wang, C.; Wang, H.; Li, J. Learn Multiple-Kernel SVMs for Domain Adaptation in Hyperspectral Data. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1224–1228.
30. Kim, W.; Crawford, M.M. Adaptive Classification for Hyperspectral Image Data Using Manifold Regularization Kernel Machines. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4110–4121. [CrossRef]
31. Bruzzone, L.; Marconcini, M. Domain Adaptation Problems: A DASVM Classification Technique and a Circular Validation Strategy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 770–787. [CrossRef] [PubMed]
32. Persello, C.; Bruzzone, L. Active Learning for Domain Adaptation in the Supervised Classification of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4468–4483. [CrossRef]
33. Persello, C. Interactive Domain Adaptation for the Classification of Remote Sensing Images Using Active Learning. *IEEE Geosci. Remote Sens. Lett.* **2012**, *10*, 736–740. [CrossRef]
34. Matasci, G.; Tuia, D.; Kanevski, M. SVM-Based Boosting of Active Learning Strategies for Efficient Domain Adaptation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1335–1343. [CrossRef]
35. Crawford, M.M.; Tuia, D.; Yang, H.L. Active Learning: Any Value for Classification of Remotely Sensed Data? *Proc. IEEE* **2013**, *101*, 593–608. [CrossRef]
36. Persello, C.; Bruzzone, L. Active and Semisupervised Learning for the Classification of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6937–6956. [CrossRef]
37. Csurka, G. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision ECCV*; Springer: Berlin, Germany, 2004; Volume 44, pp. 1–22.
38. Bosch, A.; Zisserman, A.; Muñoz, X. Scene Classification Using a Hybrid Generative/Discriminative Approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 712. [CrossRef] [PubMed]
39. Lazebnik, S.; Schmid, C.; Ponce, J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 2169–2178.
40. Ukil, A. Support Vector Machine. *Comput. Sci.* **2002**, *1*, 1–28.
41. Cheng, M.; Mitra, N.J.; Huang, X.; Torr, P.H.S. Global Contrast Based Salient Region Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 569–582. [CrossRef] [PubMed]
42. Itti, L.; Koch, C.; Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]
43. Koch, C.; Ullman, S. Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum. Neurobiol.* **1985**, *4*, 219. [PubMed]
44. Ma, Y.F.; Zhang, H.J. Contrast-based image attention analysis by using fuzzy growing. In Proceedings of the ACM International Conference on Multimedia, Berkeley, CA, USA, 2–8 November 2003; pp. 374–381.
45. Yuan, Z.; Yuan, Z.; Sun, J.; Zheng, N.; Zheng, N.; Tang, X.; Shum, H.Y. Learning to Detect a Salient Object. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 353–367.
46. Goferman, S.; Zelnik-Manor, L.; Tal, A. Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1915. [CrossRef] [PubMed]
47. Zhai, Y.; Shah, M. Visual attention detection in video sequences using spatiotemporal cues. In Proceedings of the ACM International Conference on Multimedia, Santa Barbara, CA, USA, 23–27 October 2006; pp. 815–824.
48. Achanta, R.; Hemami, S.; Estrada, F.; Susstrunk, S. Frequency-tuned salient region detection. In Proceedings of the 2009. IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1597–1604.
49. Ko, B.C.; Nam, J.Y. Object-of-interest image segmentation based on human attention and semantic region clustering. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **2006**, *23*, 2462. [CrossRef] [PubMed]
50. Han, J.; Ngan, K.N.; Li, M.; Zhang, H.J. Unsupervised extraction of visual attention objects in color images. *IEEE Trans. Circuits Syst. Video Technol.* **2006**, *16*, 141–145. [CrossRef]
51. Reinagel, P.; Zador, A.M. Natural scene statistics at the centre of gaze. *Network* **1999**, *10*, 341–350. [CrossRef] [PubMed]
52. Weijer, J.V.D.; Gevers, T.; Bagdanov, A.D. Boosting color saliency in image feature detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 150–156. [CrossRef] [PubMed]
53. Yan, L.; Zhu, R.; Mo, N.; Liu, Y. Improved Class-Specific Codebook with Two-Step Classification for Scene-Level Classification of High Resolution Remote Sensing Images. *REMOTE SENS-BASEL* **2017**, *9*, 223. [CrossRef]

54. Qi, K.; Yang, C.; Guan, Q.; Wu, H.; Gong, J. A Multiscale Deeply Described Correlatons-Based Model for Land-Use Scene Classification. *Remote Sens.* **2017**, *9*, 917. [CrossRef]
55. Zhao, L.J.; Tang, P.; Huo, L.Z. Land-Use Scene Classification Using a Concentric Circle-Structured Multiscale Bag-of-Visual-Words Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *7*, 4620–4631. [CrossRef]
56. Khan, R.; Barat, C.; Muselet, D.; Ducottet, C. Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model. *Comput. Vis. Image Underst.* **2015**, *132*, 102–112. [CrossRef]
57. Soares, C.; Brazdil, P.B.; Kuba, P. A Meta-Learning Method to Select the Kernel Width in Support Vector Regression. *Mach. Learn.* **2004**, *54*, 195–209. [CrossRef]
58. Miranda, P.B.C.D.; Prudêncio, R.B.C.; de Carvalho, A.C.P.L.; Soares, C. Combining a multi-objective optimization approach with meta-learning for SVM parameter selection. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Seoul, Korea, 14–17 October 2012; pp. 2909–2914.
59. Reif, M.; Shafait, F.; Dengel, A. Meta-learning for evolutionary parameter optimization of classifiers. *Mach. Learn.* **2012**, *87*, 357–380. [CrossRef]
60. Ali, S.; Smith-Miles, K.A. A meta-learning approach to automatic kernel selection for support vector machines. *Neurocomputing* **2006**, *70*, 173–186. [CrossRef]
61. Shafer, S.A. Using color to separate reflection components. *Color Res. Appl.* **1985**, *10*, 210–218. [CrossRef]
62. Xia, G.S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 1–17. [CrossRef]
63. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the Sigspatial International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 270–279.
64. Hu, J.; Jiang, T.; Tong, X.; Xia, G.S. A benchmark for scene classification of high spatial resolution remote sensing imagery. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 5003–5006.
65. Zhong, Y.; Zhu, Q.; Zhang, L. Scene Classification Based on the Multifeature Fusion Probabilistic Topic Model for High Spatial Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6207–6222. [CrossRef]
66. Bosch, A.; Zisserman, A.; Muñoz, X. Scene classification via Plsa. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 517–530.
67. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
68. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; Lecun, Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. *arXiv*, 2014.
69. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Darrell, T. Jonathan Caffe: Convolutional Architecture for Fast Feature Embedding. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 675–678.
70. Szegedy, 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 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
71. 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]
72. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]