Hindawi
Mathematical Problems in Engineering
Volume 2019, Article ID 7390327, 13 pages
https://doi.org/10.1155/2019/7390327

Research Article

Hybrid Relative Attributes Based on Sparse Coding for Zero-Shot Image Classification

Nannan Lu, Yanjing Sun, and Xiao Yun

School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, China

Correspondence should be addressed to Nannan Lu; lnn_921@126.com

Received 15 June 2018; Revised 23 December 2018; Accepted 7 February 2019; Published 21 February 2019

Academic Editor: Emilio Insfran Pelozo

As a specific case of image recognition, zero-shot image classification is difficult to solve since its training set cannot cover all the categories of the testing set. From the view point of human vision recognition, the objects can be recognized through the visible and nameable description to the properties. To be the semantic description of the object property, attributes can be taken as a bridge between the seen and unseen categories, which are capable of using into zero-shot image classification. There are mainly binary attributes and relative attributes for zero-shot classification, where the relative attributes have the ability to catch more general sematic relationship than the binary ones. But relative attributes do not always work in zero-shot classification for those categories having similar relative strength attributes. Aiming at solving the defect of the relative attributes in describing the similar categories, we propose to construct the Hybrid Relative Attributes based on Sparse Coding (SC-HRA). First, sparse coding is implemented on low-level features to get nonsemantic relative attributes, which are the necessary complement to the existing relative attributes. After that, they are integrated with the relative attributes to form the hybrid relative attributes (HRA). HRA ranking functions are then learned by the relative attribute learning. Finally, the class label is obtained according to the predicted ranking results of HRA and the ranking relations of HRA among the categories. To verify the effectiveness of SC-HRA, the extensive experiments are conducted on the datasets of faces and natural scenes. The results show that SC-HRA acquires the higher classification accuracy and AUC value.

1. Introduction

Image recognition has attracted the attention of many researchers and made a lot of progress in recent years. In general, the methods applied to image recognition problems include supervised learning, semisupervised learning and unsupervised learning. However, due to the high cost of the procedure of labeling images, it is not enough to cover all the categories to discriminate merely by the labeled images. Also no labeled image can be obtained. Zero-shot image classification is therefore proposed to solve the problem where the unseen categories of the training set exist in the testing set. To solve zero-shot classification problems, traditional image recognition approaches build the relations between low-level features and object category labels. However, the recent approaches prefer to use semantic attributes.

As it is well known, attribute [1] is the property that can be described by semantics or tagged by persons (e.g., with wings, black hair). Thus, attribute and feature are used to describe image contents from different levels. In low-level description of the image content, feature can be recognized by machines but cannot be understood by humans. However, attribute is the high-level description of the image content which can be understood by humans as well as machines. This is because attribute learning has shown excellent characteristic in describing objects, usability in face recognition [2], image retrieval [3, 4], image description [5, 6], and image classification [7, 8]. An object usually has various properties. For example, the hawks have wings, Asians have black hair, the road is open, and high-heeled shoes have high heel, etc. Through these attributes, people can recognize the objects' appearance and describe them in a semantic way. Besides, different categories may have the same attributes which can be regarded as the middle level description of objects in the classifier to be shared between different categories. In this way, the known knowledge can be transferred from one category to other new categories to make classifiers identify the categories even without any training samples.

The relative attribute [9], also named semantic relative attribute, indicates the attribute strength in an image compared to that in other image. The range of relative attribute is $(-\infty, +\infty)$, while the absolute value of it has no meaning because of its comparability between different categories. In such a way, relative attribute has many visual applications to help visual tasks understand the image contents as humans prefer to understand objects in the comparative way. Compared to the binary attribute, relative attribute has a richer language description for objective contents. Therefore, relative attribute has a strong ability in image description and human-computer interaction, which can effectively bridge the semantic gap between the low-level image feature and the high-level semantic feature.

Owing to the advantage of comparative description ability, relative attributes are used to zero-shot image classification [10–13]. Parikh et al. [9] demonstrated that the relative attribute learning could acquire higher classification accuracy in zero-shot image classification than the binary attributes. In relative attributes-based zero-shot classification, the relative strengths of the attributes are considered, which comparatively describe the samples in attributes like "greater than", "less than", or "equal to". By this way, the sample has distinguishable location in the attribute space, and the class label of the sample can be further determined. However, for those categories whose attributes have similar relative strengths (that is the attribute locations are close to each other in the attribute space), it is difficult to distinguish them only by the relative attributes. To solve this problem, a hybrid relative attribute model is constructed using sparse coding. Firstly, the model utilizes sparse coding to reconstruct the low-level feature of the image [14] and then obtain the low dimensional and compact representation of the image. Furthermore, the relative relations among the reconstructed features are regarded as a new kind of relative attributes. Due to its no semantic character, the reconstructed feature is named as nonsemantic relative attribute, which can be taken as the complement of the relative attributes and further increase the distinguishability of the relative attributes. Therefore, it is easy to distinguish the categories with the attributes having the similar relative relation strengths [15]. Finally, the constructed hybrid relative attributes are used to improve the performance of zero-shot image classification.

The rest of the paper is organized as follows. The related works are reviewed in Section 2. Section 3 describes the hybrid relative attribute learned by sparse coding in detail. In Section 4, the experiments are implemented on the zero-shot image classification. Finally, the conclusion is given in Section 5.

2. Related Works

Among the image recognition problems, zero-shot classification is comparably difficult to deal with. As training set and testing set have different label space, some researchers considered it as a special case of transfer learning [16] and then find the sharable low-level feature subspace by minimizing the discrimination between training set and testing set. On the other aspect, images naturally have rich linguistic meaning. For the cases without labeled training samples, the visual attributes bridge the gap between low-level features and labels. Attribute learning has excellent performance on object description [12] and knowledge transfer [17], which can be successfully applied to zero-shot image classification.

Attribute learning mainly contains binary attribute learning and relative attribute learning. The binary attribute learning focuses on those attributes with Boolean values 0 and 1. 0 denotes the absence of the attribute in an object, whereas 1 means the existence of the attribute in the object. Commonly, support vector machine [18] is used to train the binary classifier. For the images with plural attributes, the binary classifier is trained on the extracted low-level feature for each attribute, where the binary attributes could assist the classifier to semantically understand image contents. However, humans always learn about the objects in a relative way. For example, Karen and Mark are smiling and Karen is smiling more than Mark, which covey different information. Therefore, even a simple relative statement about an attribute could convey distinguishable meaning and abundant information of the image content [19]. Generally, the relative attribute mainly focuses on comparatively describing images through the relative strengths of the attribute in different samples [9]. The relative attribute has been used to search images using the feedback mechanism. Once the image search started, the relevance function was updated, and the image pool was reranked based on the accumulated constraints [20]. Moreover, the attribute feedback was used in active learning, which further prompted its classification ability [21]. Sandeep et al. [22] proposed the local parts instead of global description to jointly represent pairs of images. The part was a block around the detected point by the domain specific method. Taking "smiling" as an example, the area of mouth should be located as the significant part with important information to distinguish images since it is highly correlated with image categories. But for "eyes-open", the best local part is definitely different from "smiling". However, neither binary attributes nor relative attributes make sense all the time. Thus, they could be used together to learn the relationship between attributes and the statements in a more natural way [19].

In general, the relative attributes can describe the objects more naturally, which conform to human language. In the case of the attributes having similar relative strengths, the relative attribute learning does not work. Inspired by the landmark point of local part, it is feasible to implement sparse coding on the low-level feature to acquire more important information as the nonsemantic attributes for the zero-shot image classification. Therefore, we propose the hybrid relative attribute model to get the compact representation of images. Combining the semantic and nonsemantic attributes, the zero-shot classification is prompted to get the higher performance.

3. SC-HRA Model Based Zero-Shot Image Classification

SC-HRA model for zero-shot image classification is mainly divided into IV stages. Stage I is to construct the hybrid relative attributes. The nonsemantic attributes of training

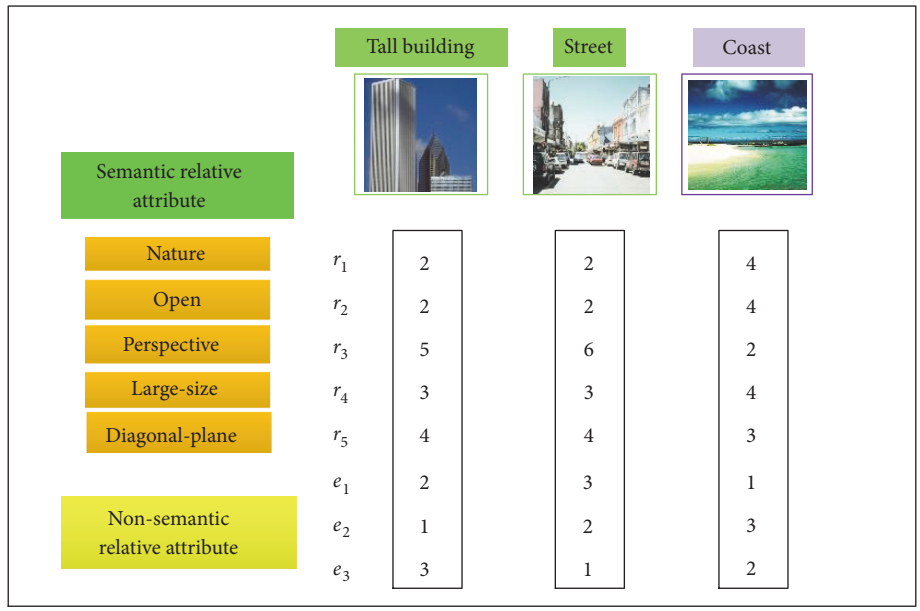


Figure 1: The hybrid relative attribute.

and testing samples are obtained by sparse coding based on the training and testing samples, respectively. And the relationship between the nonsemantic relative attribute and the class label is obtained simultaneously. Then the semantic relative attributes and the nonsemantic relative attributes are integrated as the hybrid relative attributes. Stage II is to learn the ranking function for each hybrid relative attribute. We use the ranking SVM to train the ordered pairs of the attributes as well as the unordered pairs of the attributes. Thus, each hybrid relative attribute has its own ranking function. Stage III is to build the Gaussian class model. For the training category, the ranking functions learned by stage II are used to build the Gaussian class model. But, for the testing category, the Gaussian class model is built by the relationship between the hybrid relative attributes of the training category and those of the testing category. Then, in terms of the established Gaussian class models, all categories can be located in the attribute space, which assist to distinguish the relative strength of the attributes in different categories. In Stage IV, the zero-shot image classification is implemented. The hybrid relative attributes of the testing samples are predicted by the attribute ranking functions of stage II. Then, Maximum Likelihood Estimation (MLE) is used to calculate the similarity between the hybrid relative attributes of the testing samples and the Gaussian models of each category to predict the class label of the testing sample.

In order to implement SC-HRA model based zero-shot image classification, the key points of the above five stages are to construct the hybrid relative attributes by learning the nonsemantic relative attributes and learn the ranking functions for hybrid relative attributes. The detailed realization is described in this section.

3.1. Hybrid Relative Attributes. In this section, we take the Outdoor Scene Recognition (OSR) [23] dataset as an example to explain the construction of the hybrid relative attribute. As shown in Figure 1, five semantic relative attributes $(r_1, r_2, r_3, r_4, r_5) \in R$ are given to describe 3 categories, which are "Tall building", "Street", and "Coast". "Tall building" and "Street" have similar ranks on all the semantic relative attributes except the attribute "Perspective". So the generative model [9] of "Tall building" built on the semantic attributes is extremely similar to that of "Street". Therefore, when the generative model is used to classify the zero-shot images, it is difficult to distinguish "Tall building" and "Street" by the classifier only depending on the semantic relative attributes. But for the category "Coast", the semantic relative attributes are far different from those describing "Tall building" and "Street". Thereby, the generative model of "Coast" is obviously distinguished from those of "Tall building" and "Street". Naturally, it is easy for the classifier to identify "Coast" from "Tall building" and "Street". Therefore, considering the limitation of the semantic relative attributes, the nonsemantic relative attributes $(e_1, e_2, e_3) \in E$ acquired by sparse coding are taken as the supplement of the semantic attributes to construct the hybrid relative attributes $H = \{R; E\}$, which aim at classifying those categories with high similarity. On the other hand, the auxiliary nonsemantic relative attributes cannot mislead the classifier which has obtained the correct results by the semantic relative attributes, such as "Coast".

3.2. Nonsemantic Relative Attribute Learning Based on Sparse Coding. The main idea of the nonsemantic attribute learning by sparse encoding is described in this part. At first, the features of the unlabeled training samples are used to learn a set of basis vectors. After, the features of all the training samples can be linearly represented by the learned basis vector set. Besides, the features of the testing samples can also be encoded by the learned basis vectors.

Input: the low level features x of the training samples and the testing samples, the weight attenuation coefficient λ.
Output: the non-semantic relative attributes e of the training samples and the testing samples.
(1) Step 1: bring the low level features x of the training samples into Equation (1), and initialize the basis vector φ;
(2) *Repeat;*
(3) Fix the basis vector φ in the last step, update the activation value e to minimize Equation (1).
(4) Fix the activation value e in the last step, update the basis vector φ to minimize Equation (1).
(5) Repeat the above procedure till convergence is reached, obtain the basis vector φ to represent the sample features well.
(6) Step 2: bring the basis vector φ and the low level feature x of the testing samples into Equation (2).
(7) Step 3: adjust the activation value to minimize Equation (2), the adjusted activation value e is the non-semantic relative attributes of the testing sample.

Algorithm 1: Nonsemantic relative attribute learning based on sparse coding.

$X = (x_1, x_2, \ldots, x_K) \in R^{d \times K}$ is a set of input vectors, where $x_i^{\mathrm{T}} \in R^{1 \times d}$, $i = 1, 2, \ldots, K$ denotes the ith input vector and d is its dimensions. Basis vectors $\Phi = \{\varphi_1, \varphi_2, \ldots, \varphi_N\}$ are learned by sparse coding, which are used to reconstruct the initial input features $\widetilde{X_i} \approx \sum_{j=1}^{N} e_j \varphi_j$ later. e_j is the activation value, which is named as the nonsemantic attribute in this paper. Generally, sparse coding has two steps: training and coding.

In the training stage, a set of basis vectors are generated using the training samples. Then, the linear combination of the set of basis vectors is used to express input vectors. Thus, the generation of the basis vectors can be regarded as an optimization problem, which can be described by the following objective function [24]:

$$\min_{\varphi, e} \quad \sum_{i=1}^{K} \left\| x_i - \sum_{j=1}^{N} e_{i,j} \varphi_j \right\|^2 + \lambda \sum_{i=1}^{K} \sum_{j=1}^{N} \left| e_{i,j} \right| \qquad (1)$$
$$s.t. \quad \left\| \varphi_j \right\|^2 \le c, \quad \forall j = 1, 2, \ldots, N$$

where $e_{i,j}$ is the jth sparse coding representation of the training sample x_i. λ denotes the weight attenuation coefficient. $|e_{ij}|$ can be taken as a sparse regularization term, which potentially obliges the constraint function to have the unique solution and ensures the distinct features to represent it. By solving the optimization problem in (1), we can obtain the activation value $e_{i,j}$ and the basis vector φ_j of the input feature vector x_i. When $e_{i,j}$ and φ_j vary simultaneously, the objective function is not always the convex optimization problem, whereas, with the basis vector φ_j fixed, it turns to be the convex optimization problem about the activation value $e_{i,j}$. If $e_{i,j}$ is fixed, it becomes the convex optimization problem about the basis vector φ_j. Therefore, it is feasible to implement the sparse coding as the alternately iterative optimization problem, which can fix one variable to get the solution of the other variable [25, 26]. Each iteration has two steps. The first step is to fix the basis vector φ and adjust the activation value e to minimize (1). The second step is to fix the activation value e and adjust the basis vector φ to minimize the objective constraint function. Till it is converged, the set of basis vectors can be obtained.

The coding stage is to calculate the sparse representation of a new sample. As the basis vector φ has been obtained in the training stage, in order to obtain the sparse representation of the new sample, it is just needed to adjust the activation value $e_{i,j}$ to minimize (2). Hence, the activation vector $(e_1, e_2, \ldots, e_N)$ is the sparse representation of the sample, which is the nonsemantic relative attributes as well.

$$\min_{e} \quad \sum_{i=1}^{K} \left\| x_i - \sum_{j=1}^{N} e_{i,j} \varphi_j \right\|^2 + \lambda \sum_{i=1}^{K} \sum_{j=1}^{N} \left| e_{i,j} \right| \qquad (2)$$

After that, the relative relations of N nonsemantic attributes among the different samples are determined by comparing the nonsemantic relative attributes $e = (e_1, e_2, \ldots, e_N)$ between the different samples. Algorithm 1 shows the specific procedure to obtain the nonsemantic relative attributes by sparse coding.

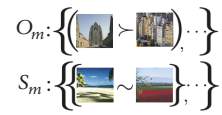


Figure 2: Ordered pairs and unordered pairs.

3.3. Ranking Function Learning Based on Hybrid Relative Attributes. The ranking functions are learned based on the hybrid relative attributes $\{h_1, h_2, \ldots, h_f\} = \{r_1, r_2, \ldots, r_M, e_1, e_2, \ldots, e_N\}$. For each hybrid relative attribute h_f ($f = 1, 2, \ldots, M + N$) of the training samples, the ordered pairs of attributes $O_f = \{(i, j)\}$ and the unordered pairs of attributes $S_f = \{(i, j)\}$ are given, respectively, where $(i, j) \in O_f \Longrightarrow i > j$ means image i has a stronger presence of attribute h_f than image j and $(i, j) \in S_f \Longrightarrow i \approx j$ means that image i and j have the similar relative strengths of h_f. As an example "open" shown in Figure 2, O_m denotes the ordered pairs. And the pair of the images above correspondingly is an example of the ordered pairs, where the left image looks opener than the right. S_m denotes the unordered pairs. The image pair below is hard to say which image is much more enlarged.

In order to obtain the relative relations between the attributes, we use Ranking SVM to learn the ranking functions $\theta_f(x_i)$ of $(M + N)$ attributes [27]:

$$\theta_f(x_i) = w_f^T x_i, \quad f = 1, 2, \ldots, M + N \qquad (3)$$

where w_f is the projection vector. For $f = 1, 2, \ldots, M+N$, the conditions shown in (4) and (5) should be satisfied:

$$\forall (i,j) \in O_f : \theta_f(x_i) > \theta_f(x_j) \tag{4}$$

$$\forall (i,j) \in S_f : \theta_f(x_i) \approx \theta_f(x_j) \tag{5}$$

Learning the ranking attribute functions is to find the optimal projection direction in the low-level feature space, so that all the samples have the right orders in the projection direction. Consequently, (4) and (5) are transformed into

$$\forall (i,j) \in O_f : w_f^T(x_i - x_j) > 0 \tag{6}$$

$$\forall (i,j) \in S_f : w_f^T(x_i - x_j) = 0 \tag{7}$$

As shown in (5) and (6), the ranking relation $i > j$ between the samples can be represented as $(i-j)$. The ordered relation between any two samples in the sharable feature space can be represented by the new vector $(x_i - x_j)$ and its corresponding new label, which is shown in

$$\{x_i - x_j, g\}_{i,j=1}^{L}, \quad g = \begin{cases} +1 & if\ i > j \\ -1 & \text{otherwise} \end{cases} \tag{8}$$

Therefore, we can correctly rank attributes by solving the standard binary classification in (8), where its optimization objective is shown as

$$w^* = \operatorname{argmin} \left(\frac{1}{2} \|w_f\|_2^2 + C \left(\sum_{i,j} \xi_{ij}^2 + \sum_{i,j} \rho_{ij}^2 \right) \right)$$

$$s.t. \quad \begin{cases} w_f^T(x_i - x_j) \geq 1 - \xi_{ij}; & \forall (i,j) \in O_f \\ \left| w_f^T(x_i - x_j) \right| \leq \rho_{ij}; & \forall (i,j) \in S_f \\ \xi_{ij} \geq 0; & \gamma_{ij} \geq 0 \end{cases} \tag{9}$$

Herein, ξ_{ij} is the nonnegative relaxation factor of the ordered pairs $O_f = \{(i,j)\}$; ρ_{ij} is the nonnegative relaxation factor of the unordered pairs $S_f = \{(i,j)\}$; and C is a trade-off constant to balance between the maximization of the margin distance and the satisfaction of the relative relation of the attribute pair. By maximizing the ranking margin $1/\|w_f\|$ and minimizing the relaxation factors ξ_{ij} and ρ_{ij} simultaneously, we can obtain the optimal projection vector w_f^*. Then, the attribute ranking function is represented as

$$\theta_f(x_i) = w_f^{*T} \cdot x_i \tag{10}$$

As to our knowledge of machine learning, binary classifier is to perfectly separate the positive and negative samples, where the margin restriction is to maximize the distance of the nearest two samples from the two different categories, whereas learning the ranking functions is to rank the training samples in terms of the attribute strengths, so its margin restriction is to maximize the distance of the nearest two samples during the ranking [28, 29]. Therefore, the ranking functions are better to reflect the relationships of the relative strengths between the attributes of the different categories.

3.4. Zero-Shot Image Classification Based on SC-HRA. Due to the zero-shot learning without known samples, the classification model is built based on the relative relations of the attributes between the training and testing samples by the method in [9]. Specifically, the images are given from Q categories, of which K categories are the "seen" categories as the training samples to learn the ranking functions of the hybrid relative attributes, and $L = Q - K$ categories are the "unseen" categories as the testing samples not involved in learning the ranking functions. In the modelling procedure, $(\mu^{(s)}, \Sigma_i^{(s)})$ and $(\mu^{(u)}, \Sigma_i^{(u)})$ are used to represent the models of the training category and the testing category, respectively.

The procedure of zero-shot image classification is explained in the following.

(1) Building the model of the training category: first, the attribute ranking scores $(\theta_i^1, \theta_i^2, \ldots, \theta_i^f, \ldots, \theta_i^{M+N})$ of the training samples are predicted by the ranking functions of the hybrid relative attributes, in which θ_i^f is the ranking score of the fth hybrid relative attribute in the ith training sample. After that, based on the ranking scores of the hybrid relative attributes in the training samples, the mean value $\mu_i^{(s)} \in R^{M+N}$ and the covariance matrix $\Sigma_i^{(s)}$ are estimated for the training categories. Finally, the Gaussian model of the training category is obtained, which is $c_i^{(s)} \sim N(\mu_i^{(s)}, \Sigma_i^{(s)})$, $i = 1, \ldots, K$.

(2) Building the model of the testing category: for the hybrid relative attribute h_f, the testing category c_j^u can be expressed into three cases based on the "seen" $c_i^{(s)}$ and $c_k^{(s)}$:

① in the case of $c_i^{(s)} > c_j^{(u)} > c_k^{(s)}$, the mean value of the model is $\mu_{f,j}^{(u)} = (1/2)(\mu_{f,i}^{(s)} + \mu_{f,\mathrm{k}}^{(s)})$, and the covariance of the model is $\Sigma_j^{(u)} = (1/K)\sum_{i=1}^{K} \Sigma_i^{(s)}$; ② in the case of $c_i^{(s)} > c_j^{(u)}$, the mean value of the model is $\mu_{f,j}^{(u)} = \mu_{f,i}^{(s)} - d_{f,\mathrm{i}}^{(s)}$, and the covariance of the model is $\Sigma_j^{(u)} = (1/K)\sum_{i=1}^{K} \Sigma_i^{(s)}$; ③in the case of $c_j^{(u)} > c_k^{(s)}$, the mean value of the model is $\mu_{f,j}^{(u)} = \mu_{f,k}^{(s)} + d_{f,j}^{(s)}$, and the covariance of the model is $\Sigma_j^{(u)} = (1/K)\sum_{i=1}^{K} \Sigma_i^{(s)}$.

From the above, $\mu_{f,i}^{(s)} = (1/L)\sum_{i=1}^{L} \theta_{f,i}^{(s)}$ represents the mean value of the fth hybrid relative attribute, and $d_{f,i}^{(s)} = (1/L)\sum_{i=1}^{L}(\theta_{f,i}^{(s)} - \mu_{f,i}^{(s)})$ represents the average difference among the ranking scores of the hybrid relative attributes in the training category. It is reasonable to assume that the relation distributions of the hybrid relative attributes of all the categories are same.

In the testing, we first compute the attribute ranking score θ_i of the ith testing sample by the ranking function of the hybrid relative attribute. Then, the class label is given by the similarity between θ_i and the category models, which is shown as

$$c^* = \arg\min_{j \in \{1,\ldots,N\}} p(\theta_i \mid \mu_j, \Sigma_j) \tag{11}$$

We also concludes the implementation of zero-shot image classification based on SC-HRA, which is shown as Algorithm 2.

Input: the attribute ranking scores of the training and testing samples
Output: the label of the testing sample
(1) *Building of the training models;*
(2) Calculate mean value $\mu_i^{(s)}$ and covariance matrix $\Sigma_i^{(s)}$ based on θ_i^f of training sample, obtain $c_i^{(s)} \sim N(\mu_i^{(s)}, \Sigma_i^{(s)})$
(3) *Building of the testing models;*
(4) If $c_i^{(s)} > c_j^{(u)} > c_k^{(s)}$ is satisfied, then the mean value of model is $\mu_{f,j}^{(u)} = (1/2)(\mu_{f,i}^{(s)} + \mu_{f,k}^{(s)})$, and the covariance matrix is $\Sigma_j^{(s)} = (1/K) \sum_{i=1}^{K} \Sigma_i^{(s)}$.
(5) If $c_i^{(s)} > c_j^{(u)}$ is satisfied, then the mean value of model is $\mu_{f,j}^{(u)} = \mu_{f,i}^{(s)} - d_{f,i}^{(s)}$, and the covariance matrix is $\Sigma_j^{(u)} = (1/K) \sum_{i=1}^{K} \Sigma_i^{(s)}$.
(6) If $c_j^{(u)} > c_k^{(s)}$is satisfied, then the mean value of model is $\mu_{f,j}^{(u)} = \mu_{f,k}^{(s)} + d_{f,j}^{(s)}$, and the covariance matrix is $\Sigma_j^{(u)} = (1/K) \sum_{i=1}^{K} \Sigma_i^{(s)}$.
(7) *Testing;*
(8) By calculating the attribute ranking score θ_i of the testing sample, the class label is determined as $c^* = \arg\min_{j \in \{1,\ldots,N\}} p(\theta_i \mid \mu_j, \Sigma_j)$

Algorithm 2: Zero-shot image classification based on SC-HRA.

Table 1: Description of the relative attributes of OSR dataset.

Attribute name	Relative attribute
natural	T<I~S<H<C~O~M~F
open	T~F<I~S<M<H~C~O
perspective	O<C<M~F<H<I<S<T
Large-objects	F<O~M<I~S<H~C<T
Diagonal-plane	F<O~M<C<I~S<H<T
Depth-close	C<M<O<T~I~S~H~F

Table 2: Description of the relative attributes of Pub Fig dataset.

Attribute name	Relative attribute
Masculine-looking	S<M<Z<V<J<A<H<C
White	A<C<H<Z<J<S<M<V
Young	V<H<C<J<A<S<Z<M
Smiling	J<V<H<A~C<S~Z<M
Chubby	V<J<H<C<Z<M<S<A
Visible-forehead	J<Z<M<S<A~C~H~V
Bushy-eyebrows	M<S<Z<V<H<A<C<J
Narrow-eyes	M<J<S<A<H<C<V<Z
Pointy-nose	A<C<J~M~V<S<Z<H
Big-lips	H<J<V<Z<C<M<A<S
Round-face	H<V<J<C<Z<A<S<M

4. Experiments

4.1. Experimental Datasets. In this section, the proposed method is evaluated on OSR dataset and Public Figure Face (Pub Fig) [9, 30] dataset. The two datasets are from the distinctive areas of landscape and human faces, which can sufficiently validate the effects of the proposed method. OSR dataset contains 2688 images, 6 attributes, and 8 scenes (tall building (T), inside-city (I), street (S), highway (H), coast (C), Open-country (O), mountain (M), and forest (F)), and use 512-dimensional gist [31] descriptor to represent the images. In the paper, 2648 images of images are selected as testing set. Pub Fig dataset consist of 772 images with 11 semantic attributes from 8 identities (Alex (A), Clive (C), Hugh (H), Jared (J), Miley (M), Scarlett (S), Viggo (V), and Zac (Z)), which is 512-dimensional gist descriptor and 30-dimensional global color features extracted from each image [32]. 560 images from Pub Fig dataset are selected as the testing set. According to the empirical values, the parameters λ and c are set as 0.5 and 1, respectively.

Tables 1 and 2 show the relative attribute rankings and the description of OSR and Pub Fig datasets, respectively. The relative attributes can be ranked by the relative strengths.

4.2. Attribute Ranking Prediction. To verify the effects of the hybrid attributes on the ranking performance, we predict the ranking scores of 2648 testing images from OSR dataset and 560 testing images from Pub Fig dataset. For each pair (i, j) of one attribute, we calculate $\theta_f(x_i)$ to get the strengths of attributes. $i > j$ means that the strength of i is higher than that of j, while $i < j$ denotes the strength of i is smaller than j. In this section, 6 learning methods are compared with the proposed method, which are B-SVM [33], RFUA [34], HAP [35], CSHAP_H, CSHAP_G [35], and RA [9]. The attribute rankings predicted by the different methods are compared with the truly relative relations of the samples. Based on the comparisons of the predicted relations with the real ones, the attributes' ranking accuracy can be calculated.

Tables 3 and 4 illustrate the performance of attribute ranking prediction by different methods. Table 3 shows the comparisons of attributes' ranking accuracy on OSR dataset. Specifically, SC-HRA obtains the best ranking accuracy for all the attributes. Table 4 shows the comparisons of attributes' ranking accuracy on Pub Fig dataset. The attribute ranking accuracy on Pub Fig is higher than other methods except for the attributes "young" and "Visible Forehead", which are not as good as than RA but better than others. Moreover, the average accuracy of ranking prediction is 89.15% and 82.29%,

Table 3: Comparisons of attributes' ranking accuracy on OSR dataset (%).

	B-SVM	RFUA	HAP	$CSHAP_H$	$CSHAP_G$	RA	SC-HRA
Natural	88.23	86.66	87.04	88.12	87.72	90.68	*91.94*
Open	87.15	83.22	83.6	84.68	83.38	87.33	*89.78*
Perspective	80.02	83.42	83.26	84.7	83.58	86.54	*90.02*
Large	77.28	81.1	80.49	82.47	81.26	85.21	*86.43*
Diagonal	63.43	83.94	83.87	84.95	84.10	87.06	*90.33*
Close	85.29	80.05	81.29	82.01	81.61	84.57	*86.41*
Average	*80.23*	*83.065*	*83.26*	*84.49*	*83.61*	*86.90*	*89.15*

Table 4: Comparisons of attributes' ranking accuracy for Pub Fig dataset (%).

	B-SVM	RFUA	HAP	$CSHAP_H$	$CSHAP_G$	RA	SC-HRA
Male	80.41	80.92	81.6	82.91	81.39	85.10	*90.34*
White	71.03	76.93	77.91	78.6	77.62	80.25	*88.72*
Young	74.62	78.45	79.42	80.72	79.47	82.56	81.15
Smiling	70.59	71.48	72.46	73.77	72.97	75.96	*80.75*
Chubby	58.45	68.78	69.71	69.03	69.5	71.49	*75.92*
Visible Forehead	68.92	84.92	85.28	87.48	86.15	90.03	87.76
Bushy Eyebrows	63.22	67.72	66.46	67.67	66.97	69.95	*74.57*
Narrow Eyes	80.53	77.82	78.82	79.59	78.61	81.78	*84.16*
Pointy Nose	70.19	70.29	71.28	73.13	72.15	75.32	*78.35*
Big Lips	68.97	71.55	72.52	72.83	72.58	74.67	*78.89*
Round Face	75.56	74.12	75.11	76.42	75.44	78.61	*84.54*
Average	*71.14*	*74.82*	*75.51*	*76.56*	*75.71*	*78.70*	*82.29*

respectively. As compared to the performance of other methods, SC-HRA gets the best performance on attribute ranking prediction, which confirms the advantage of HRA during the attribute ranking.

4.3. Zero-Shot Image Classification Results. In this part, the effects of SC-HRA on the zero-shot image classification are checked. In each dataset, a series of experiments are carried out on zero-shot image classification. And for each experiment, selected are the different training ("seen") and testing ("unseen") categories to realize the classification. We adopt C_T^t fold cross-validation (T is the total number of the dataset categories, and t is the number of the training categories) to eliminate the influence of the experiment randomness. So, the results based on the cross-validation are reliable since all the samples are almost considered during training. Specifically, 5 cross-validation experiments are implemented on the bench mark datasets, which are $C_8^6 = 28$ fold, $C_8^5 = 56$ fold, $C_8^4 = 70$ fold, $C_8^3 = 56$ fold, and $C_8^5 = 28$ fold.

Figure 3 shows the results of the different number of the training categories t as we increase the nonsemantic relative attribute dimension N, where N ranges from 0 to 14. In general, the classification accuracy declines with the training categories t decreasing. The reason is that, when the number of the training categories t reduces, the fewer ordered pairs and unordered pairs are used to learn the ranking functions of the hybrid relative attributes, which impacts the ranking accuracy of the hybrid relative attributes. Thus, too low proportion of the training category can result in the worse performance of zero-shot classification. Besides, with the training categories reducing, the proportion of the categories to build "unseen" model also decreases, which is another factor for the worse performance of the classifier. Without nonsemantic relative attribute (that is, $N = 0$), HRA and RA are equivalent.

An additional phenomenon in Figure 3 is that all the curves grow first and then decline. And in the condition of t=6, the proposed method obtains the better performance on both OSR dataset and Pub Fig dataset than the other cases (t=2, t=3, t=4, t=5). Next, we calculate the average accuracy of zero-shot classification shown as Tables 5 and 6 to concretely analyse that fact. With the increase of the nonsemantic relative attribute dimension N, the average classification accuracy on OSR dataset grows to 48.02% at $N = 10$ firstly and then declines. On the other hand, the average classification accuracy of Pub Fig dataset reaches to 47.33% at N=6 and then reduces, too. As to the reason, the nonsemantic relative attributes contribute to the difference of the relative attribute space, which impacts on the classification accuracy. However, due to the increase of the nonsemantic relative attribute dimension N, too many nonsemantic relative attributes make the relative attributes redundant, which results in the decrease of the classification accuracy. Therefore, according the analysis based on the experiments, parameter N is set as 10 for OSR

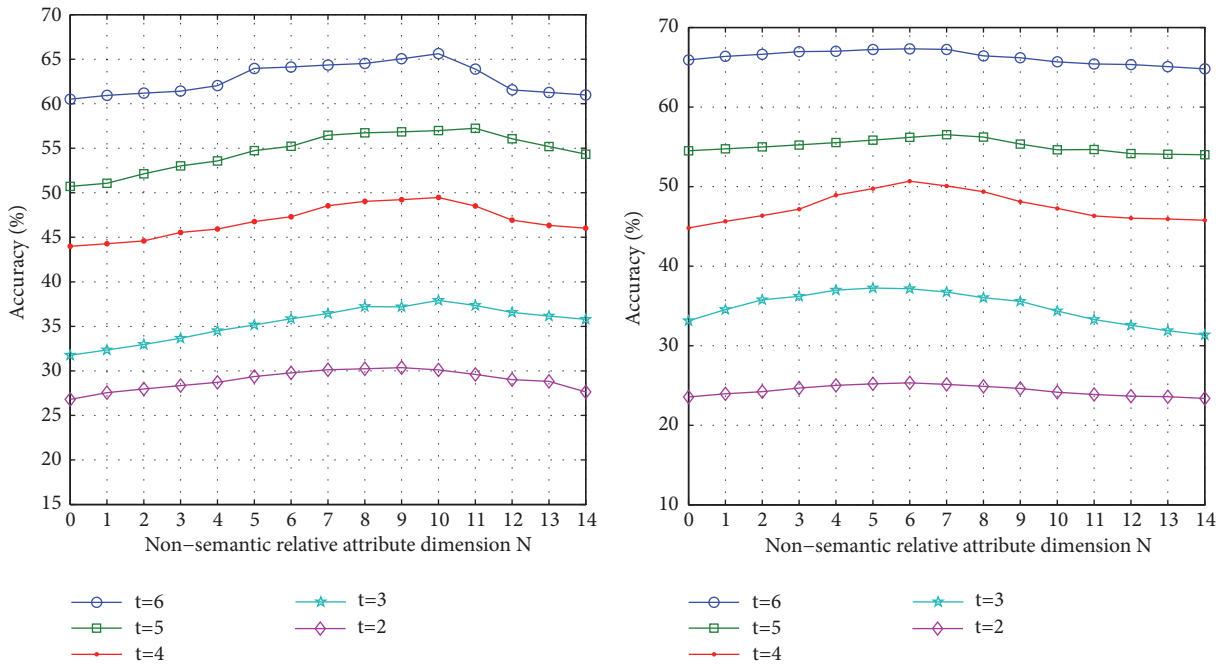


Figure 3: Zero-shot learning performance of training categories t as the nonsemantic relative attribute dimension N increases.

Table 5: Average accuracy of zero-shot image classification on OSR dataset.

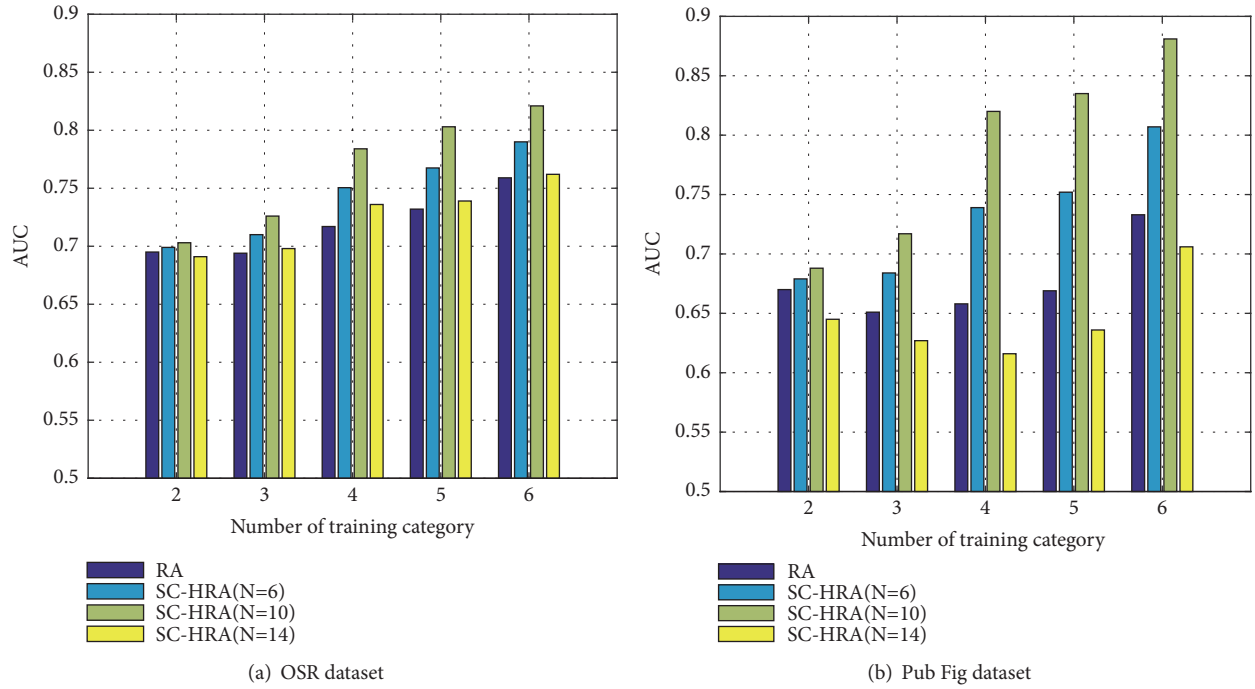
Model		Average Accuracy (%)
RA		42.75
SC-HRA	N=1	43.24
	N=2	43.77
	N=3	44.40
	N=4	44.95
	N=5	46.00
	N=6	46.46
	N=7	47.19
	N=8	47.55
	N=9	47.73
	N=10	*48.02*
	N=11	47.32
	N=12	46.03
	N=13	45.56
	N=14	44.95

Table 6: Average accuracy of zero-shot image classification on Pub Fig dataset.

Model		Average Accuracy (%)
RA		44.38
SC-HRA	N=1	45.05
	N=2	45.59
	N=3	46.04
	N=4	46.69
	N=5	47.05
	N=6	*47.33*
	N=7	47.14
	N=8	46.59
	N=9	45.97
	N=10	45.21
	N=11	44.71
	N=12	44.35
	N=13	44.11
	N=14	43.86

dataset and 6 for Pub Fig dataset when t equals 6, which can acquire the highest average accuracy.

Although the proportion of the correct classified samples to the whole testing set can be illustrated by the classification accuracy, it neglects the relation between false positive rate (the probability of negative samples wrongly categorized as positive) and true positive rate (the probability of positive samples correctly categorized as negative). Therefore, we further use AUC (Area Under Curve) [36, 37] to evaluate the classification performance. AUC denotes the area under the receiver operating characteristic (ROC) curve which provides another way to evaluate the performance of the proposed model. If the model is ideal, its AUC value equals 1. The AUC value of a random model equals 0.5. On the other hand, regarding the results of Tables 5 and 6, SC-HRA acquires the highest classification accuracy on OSR dataset at N=10.



(a) OSR dataset (b) Pub Fig dataset

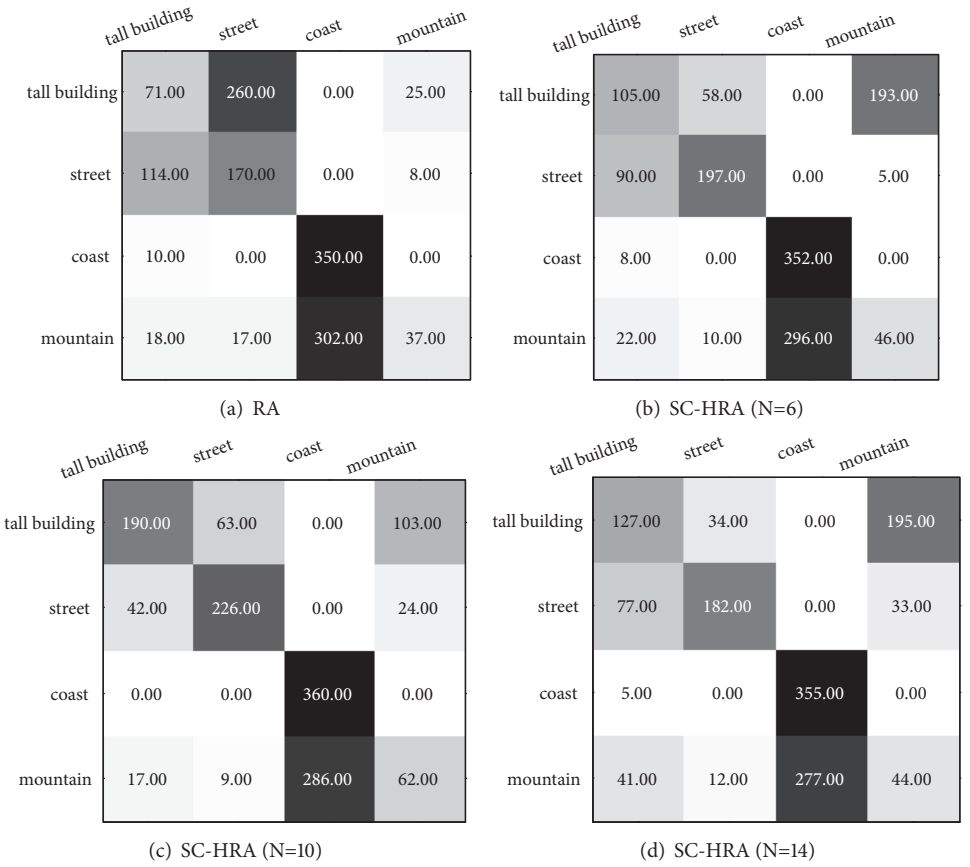
Figure 4: AUC value of zero-shot image classification.

When N equals 6, SC-HRA acquires the highest average classification accuracy on Pub Fig dataset. When N rises to 14, the classification accuracy of SC-HRA on both OSR and Pub Fig is lower than that of SC-HRA with other 13 different N values. Even worse, it is lower than that of RA. Thus, we choose the representative results of SC-HRA as well as RA, which are N=6, N=10, and N=14. Figure 4 shows the AUC values of zero-shot classification on OSR and Pub Fig datasets. In general, all the three cases of SC-HRA (N=6, N=10, and N=14) illustrated in Figure 4 have higher AUC values than the random tested AUC value (0.5) on OSR dataset as well as Pub Fig dataset. Specifically, as to N=6 and N=10, AUC values of SC-HRA are higher than RA. But as to N=14, the AUC value decreases since the redundant nonsemantic relative attributes deteriorate the performance of zero-shot image classification.

Next, we demonstrate the proposed method using the confusion matrix. Figures 5 and 6 show the classification performance of RA and SC-HRA on OSR and Pub Fig illustrated by confusion matrix. In the confusion matrix, the blocks distributed on the diagonal line denote the number of the samples correctly classified. The darker the color of the block, the larger the number of the correctly classified samples. As shown in Figure 5, due to the high similarity of the semantic relative attributes between Tall building and Street, it is difficult for RA to distinguish them. For example, 260 images of Tall building are mistaken as Street category. When the nonsemantic relative attributes are introduced into zero-shot classification, SC-HRA can distinguish Tall building and Street partially, which is better than RA. Concluded from Figures 5 and 6, in the case of $N = 6$ and $N = 10$, the samples correctly classified by SC-HRA are more than those correctly classified by RA on the diagonal lines of OSR and Pub Fig confusion matrix. The reason is that a certain amount of nonsemantic relative attributes can improve the performance of the classifier. In the case of $N = 14$, the number of samples correctly classified by SC-HRA decreases, which is caused by the redundant nonsemantic relative attributes. Therefore, a certain amount of nonsemantic relative attributes can benefit zero-shot classification to ameliorate its performance, yet too many of them bring much redundant information that may confuse the classification procedure.

4.4. Comparisons of Zero-Shot Image Classification Methods. To further verify the classification performance of the proposed method, we compare the zero-shot image classification accuracy of it with the other 6 methods, which are DAP [38], RFUA-SRA, HAP-SRA, CSHAP$_H$-SRA, CSHAP$_G$-SRA, and MLE-RA. The baseline DAP is a direct attribute prediction model. Based on RA approach, RFUA-SRA (RFUA score-based relative attributes) replaces rank values with the RFUA classifier output score; HAP-SRA (HAP score-based relative attributes) replaces rank values with the HAP classifier output score; CSHAP$_H$-SRA (CSHAP$_H$ score-based relative attributes) replaces rank values with the RFUA classifier output score; and CSHAP$_G$-SRA (CSHAP$_G$ score-based relative attributes) replaces rank values with the CSHAP$_G$ classifier output score. MLE-RA utilizes maximum likelihood estimation to predict the testing sample label.

To avoid the randomness during the simulations, the cross-validation is adopted to compare the proposed model with other baseline models. The comparisons are illustrated as Tables 7 and 8. As the proportion of training category decreases, the classification accuracies of all the models



(a) RA (b) SC-HRA (N=6)

(c) SC-HRA (N=10) (d) SC-HRA (N=14)

Figure 5: Classification results on OSR dataset represented as confusion matrix.

Table 7: Comparisons of zero-shot image classification accuracies on OSR dataset (%).

Taining categories	DAP	RFUA-SRA	HAP-SRA	$CSHAP_H$-SRA	$CSHAP_G$-SRA	MLE-RA	SC-HRA
6	54.15	55.29	56.87	58.16	57.11	60.50	62.92
5	37.64	44.41	46.99	48.28	47.32	50.71	54.97
4	27.37	38.14	40.63	42.01	40.69	43.99	47.04
3	24.48	27.02	27.69	29.87	28.29	31.76	35.65
2	20.80	22.30	23.52	24.90	23.58	26.79	29.13
Average	*32.888*	*37.432*	*39.14*	*40.644*	*39.40*	*42.75*	*45.94*

decline. Thus, each model can obtain the highest accuracy when the number of training category is 6. Further, as we expected, our method significantly outperforms DAP, RFUA-SRA, HAP-SRA, $CSHAP_H$-SRA, and $CSHAP_G$-SRA and slightly exceeds the baseline model MLE-RA. This validates the power of the hybrid relative attribute jointly combined semantic relative attribute and nonsemantic attribute for zero-shot image classification.

5. Conclusions

Zero-shot image classification primarily solves the challenge of how to identify the "unseen" categories accurately when the labeled samples cannot cover all the objective categories. In that the semantic relative attributes become the key to solve zero-shot image classification. However, the existing zero-shot image classifiers based on semantic relative attributes ignore the low-level features. In other words, the class labels of the samples in the testing categories are decided only depending on the semantic relative attributes of the samples. But for the categories having similar semantic relative attributes, it is difficult to distinguish them. Therefore, considering the restriction of the semantic relative attributes, the nonsemantic relative attributes are obtained by sparse coding the low-level features of the samples and treated as the supplement of the semantic relative attributes to construct the hybrid relative attributes. Further, the hybrid relative attributes are applied to zero-shot image classification. So a zero-shot image classification based on the hybrid relative attributes is proposed. The proposed SC-HRA is demonstrated and

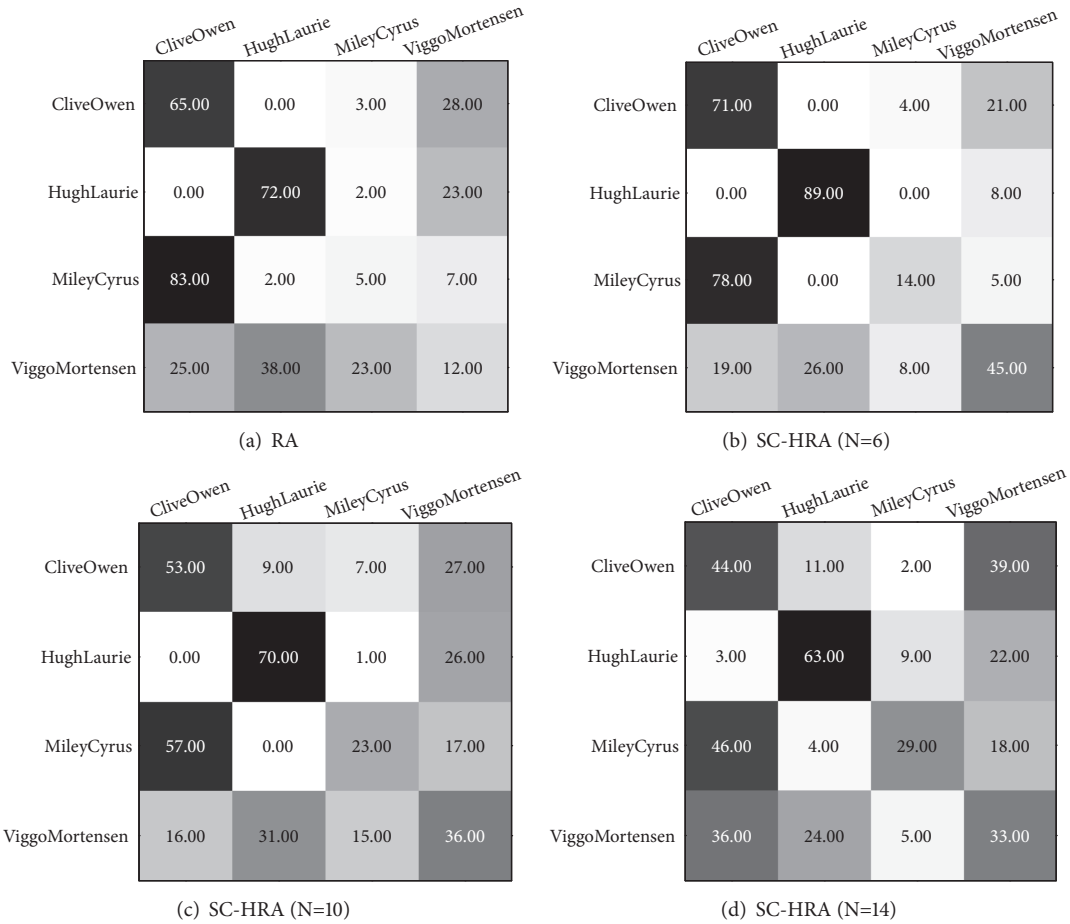


Figure 6: Classification results on Pub Fig dataset represented as confusion matrix.

Table 8: Comparisons of zero-shot image classification accuracies on Pub Fig dataset (%).

Taining categories	DAP	RFUA-SRA	HAP-SRA	$CSHAP_H$-SRA	$CSHAP_G$-SRA	MLE-RA	SC-HRA
6	63.40	61.57	62.83	63.72	62.57	65.92	66.24
5	46.91	48.15	49.49	51.48	50.12	54.50	55.12
4	37.18	40.60	41.24	42.31	41.43	44.80	47.47
3	21.01	28.44	29.62	30.48	29.65	33.13	34.85
2	16.80	19.94	20.58	20.69	20.77	23.54	24.35
Average	*37.06*	*39.74*	*40.75*	*41.74*	*40.91*	*44.38*	*45.61*

compared with the baseline methods by the experiments on OSR and Pub Fig datasets. The experimental results validate that SC-HRA has powerful ranking and classification ability in zero-shot learning due to the distinguishability of the nonsemantic relative attributes in classification.

Data Availability

All the experiments of this paper are based on the datasets of Outdoor Scene Recognition (OSR) and a subset of the Public Figure Face Database (Pub Fig), which can be downloaded from "https://www.cc.gatech.edu/~parikh/relative.html#data".

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Chinese Universities Scientific Fund (Grant no. 2015QNB20).

References

[1] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorizationa," *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 3, pp. 453–465, 2014.

[2] N. Kumar, A. Berg, P. N. Belhumeur, and S. Nayar, "Describable visual attributes for face verification and image search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1962–1977, 2011.

[3] M. Douze, A. Ramisa, and C. Schmid, "Combining attributes and Fisher vectors for efficient image retrieval," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011*, pp. 745–752, USA, June 2011.

[4] B. Siddiquie, R. S. Feris, and L. S. Davis, "Image ranking and retrieval based on multi-attribute queries," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011*, pp. 801–808, June 2011.

[5] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1778–1785, Miami, Fla, USA, June 2009.

[6] D. Mahajan, S. Sellamanickam, and V. Nair, "A joint learning framework for attribute models and object descriptions," in *Proceedings of the 2011 IEEE International Conference on Computer Vision, ICCV 2011*, pp. 1227–1234, Spain, November 2011.

[7] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid, "Label-embedding for attribute-based classification," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2013*, pp. 819–826, USA, June 2013.

[8] Y. Xian, Z. Akata, G. Sharma, Q. Nguyen, M. Hein, and B. Schiele, "Latent embeddings for zero-shot classification," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 69–77, Las Vegas, Nev, USA, June 2016.

[9] D. Parikh and K. Grauman, "Relative attributes," in *Proceedings of the 2011 IEEE International Conference on Computer Vision, ICCV 2011*, pp. 503–510, Spain, November 2011.

[10] S. Chaudhuri, E. Kalogerakis, S. Giguere, and T. Funkhouser, "Attriblt: content creation with semantic attributes," in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST 2013*, pp. 193–202, October 2013.

[11] B. Hariharan, S. V. Vishwanathan, and M. Varma, "Efficient max-margin multi-label classification with applications to zero-shot learning," *Machine Learning*, vol. 88, no. 1-2, pp. 127–155, 2012.

[12] M. Elhoseiny, B. Saleh, and A. Elgammal, "Write a classifier: zero-shot learning using purely textual descriptions," in *Proceedings of the 2013 14th IEEE International Conference on Computer Vision, ICCV 2013*, pp. 2584–2591, Australia, December 2013.

[13] M. Liu, D. Zhang, and S. Chen, "Attribute relation learning for zero-shot classification," *Neurocomputing*, vol. 139, pp. 34–46, 2014.

[14] G. Shi, D. Liu, D. Gao et al., "Advances in theory and application of compressed sensing," *Acta Electronica Sinica*, vol. 37, no. 5, pp. 1070–1081, 2009.

[15] V. Sharmanska, N. Quadrianto, and C. H. Lampert, "Augmented attribute representations," in *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 7576 of *Lecture Notes in Computer Science*, pp. 242–255, Springer, Berlin, Germany, October 2012.

[16] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Unsupervised domain adaptation for zero-shot learning," in *Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV 2015*, pp. 2452–2460, December 2015.

[17] G. Christie, A. Parkash, U. Krothapalli, and D. Parikh, "Predicting user annoyance using visual attributes," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 3630–3637, June 2014.

[18] P. Bouboulis, S. Theodoridis, C. Mavroforakis, and L. Evaggelatou-Dalla, "Complex support vector machines for regression and quaternary classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 6, pp. 1260–1274, 2015.

[19] A. Sadovnik, A. Gallagher, D. Parikh, and T. Chen, "Spoken attributes: mixing binary and relative attributes to say the right thing," in *Proceedings of the 2013 14th IEEE International Conference on Computer Vision, ICCV 2013*, pp. 2160–2167, December 2013.

[20] A. Kovashka, D. Parikh, and K. Grauman, "Whittlesearch: image search with relative attribute feedback," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2973–2980, June 2012.

[21] A. Biswas and D. Parikh, "Simultaneous active learning of classifiers & attributes via relative feedback," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2013*, pp. 644–651, June 2013.

[22] R. N. Sandeep, Y. Verma, and C. V. Jawahar, "Relative parts: distinctive parts for learning relative attributes," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 3614–3621, USA, June 2014.

[23] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[24] X.-Z. Qi and Q. Wang, "An image classification approach based on sparse coding and multiple kernel learning," *Acta Electronica Sinica*, vol. 40, no. 4, pp. 773–779, 2012.

[25] L.-L. Huang, L. Xiao, and Z.-H. Wei, "A nonlocal sparse representation method for color demosaicking," *Tien Tzu Hsueh Pao/Acta Electronica Sinica*, vol. 42, no. 2, pp. 272–279, 2014.

[26] Y. Sun, Z. Wei, L. Xiao et al., "Multimorphology sparsity regularized image super-resolution," *Acta Electronica Sinica*, vol. 38, no. 12, pp. 2898–2903, 2010.

[27] S. Wang, D. Tao, and J. Yang, "Relative attribute SVM+ learning for age estimation," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 827–839, 2016.

[28] D. Tao, L. Jin, Y. Yuan, and Y. Xue, "Ensemble manifold rank preserving for acceleration-based human activity recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1392–1404, 2016.

[29] O. C. Hamsici and A. M. Martinez, "Multiple ordinal regression by maximizing the sum of margins," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 10, pp. 2072–2083, 2016.

[30] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 365–372, IEEE, Kyoto, Japan, October 2009.

[31] T. L. Berg, A. C. Berg, and J. Shih, "Automatic attribute discovery and characterization from noisy web data," in *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 6311, pp. 663–676, Springer, Berlin, Germany, September 2010.

[32] J. Chu and G.-H. Zhao, "Scene classification based on SIFT combined with GIST," in *Proceedings of the 2014 International*

Conference on Information Science, Electronics and Electrical Engineering, ISEEE 2014, pp. 331–336, April 2014.

[33] S. Li, S. Shan, and X. Chen, "Relative forest for attribute prediction," in *Proceedings of the Asian Conference on Computer Vision*, pp. 316–327, November 2012.

[34] D. Jayaraman and K. Grauman, "Zero-shot recognition with unreliable attributes," in *Proceedings of the 28th Annual Conference on Neural Information Processing Systems 2014, NIPS 2014*, pp. 3464–3472, December 2014.

[35] S. Huang, M. Elhoseiny, A. Elgammal, and D. Yang, "Learning hypergraph-regularized attribute predictors," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 409–417, Boston, Mass, USA, June 2015.

[36] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[37] W.-H. Lee, P. D. Gader, and J. N. Wilson, "Optimizing the area under a receiver operating characteristic curve with application to landmine detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 2, pp. 389–397, 2007.

[38] C. L. Castro and A. P. Braga, "Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 6, pp. 888–899, 2013.