information

MDPI

Article

Image Aesthetic Assessment Based on Latent Semantic Features

Gang Yan, Rongjia Bi, Yingchun Guo * and Weifeng Peng

School of Artificial Intelligence, Hebei University of Technology, Tianjin 300400, China;
yangang@hebut.edu.cn (G.Y.); 201722102032@stu.hebut.edu.cn (R.B.); 201522101004@stu.hebut.edu.cn (W.P.)
* Correspondence: gyc@scse.hebut.edu.cn

Received: 29 February 2020; Accepted: 15 April 2020; Published: 17 April 2020

Abstract: Image aesthetic evaluation refers to the subjective aesthetic evaluation of images. Computational aesthetics has been widely concerned due to the limitations of subjective evaluation. Aiming at the problem that the existing evaluation methods of image aesthetic quality only extract the low-level features of images and they have a low correlation with human subjective perception, this paper proposes an aesthetic evaluation model based on latent semantic features. The aesthetic features of images are extracted by superpixel segmentation that is based on weighted density POI (Point of Interest), which includes semantic features, texture features, and color features. These features are mapped to feature words by LLC (Locality-constrained Linear Coding) and, furthermore, latent semantic features are extracted using the LDA (Latent Dirichlet Allocation). Finally, the SVM classifier is used to establish the classification prediction model of image aesthetics. The experimental results on the AVA dataset show that the feature coding based on latent semantics proposed in this paper improves the adaptability of the image aesthetic prediction model, and the correlation with human subjective perception reaches 83.75%.

Keywords: Image aesthetics assessment; density weighted Point of Interest; local constraint linear coding; latent semantic feature

1. Introduction

With the development of mobile internet technology, there are millions of online pictures uploaded and shared every day. Additionally, the need for automatic management of large-scale image datasets is growing [1]. People like to collect and save high-quality pictures [2]. In recent years, the image quality assessment from the aesthetic point of view has received extensive attention [3–7]. An automatic assessment of image aesthetics [8] has been widely used in image retrieval [9], image cropping and repair [10,11], visual design [12,13], and so on.

However, in the field of computer vision and image processing, it is a very challenging task to automatically realize the image quality by machines. There are no uniform measurement and assessment rules for the aesthetic assessment of images, and it is very subjective to judge whether a picture has high aesthetics perception or not, because each person's experience and the accumulation of aesthetic experience is different, and the judgment for the same picture is different. Computable aesthetic assessment is currently the main research method, and its research purpose is to enable computers to automatically perform quantitative analysis, calculation, and assessment of images. Researchers have established preliminary aesthetic assessment frameworks and standards, such as the composition rules in photography art and the theory of color application in painting, through the understanding and researching of image aesthetics. Although people's judgments on aesthetics have subjective factors, human aesthetic experience and habits have some common characteristics, which provide direction for researchers to use machine learning methods to simulate human aesthetic perception to make aesthetic decisions. Additionally, using the method of machine learning, through training and learning, the evaluation model of image aesthetics can be established, and the machine can automatically evaluate the aesthetic value of image and make the aesthetic evaluation similar to human beings.

This paper proposes a new aesthetic assessment model. Firstly, the image aesthetic features are extracted, which include semantic features, texture features, color features, and the extracted features are adaptively weighted according to the density of POIs (Point of Interest) in the image superpixel block. Subsequently, two coding schemes are performed on the extracted features: the first scheme is the LLC coding, which encodes the extracted features into semantic features containing image aesthetic information; the second scheme is to further quantize the extracted features, map them into feature words, and utilize the LDA (Latent Dirichlet Allocation) model [14] to extract the common features of aesthetic images—latent semantic features; finally, in the AVA dataset [15], the machine learning algorithm is used to perform the feature combination optimization, and the optimal aesthetic evaluation classification model is obtained.

The main contribution of this paper are as follows:

(1) The superpixel block algorithm that is based on the weighted density of POI is designed to extract local handcraft features of the image. The spatial information loss and the feature dimension are both dramatically reduced, due to introducing spatial attributes. Additionally, the density of POI measures the complexity of local areas. It increases the relationship between image features and aesthetic complexity attributes.

(2) The aesthetic features are coded by visual word and then mapped into multiple semantic documents. Additionally, the indescribable aesthetic information is mined through the feature documents. The combination of invisible latent semantic features and visible semantic features greatly improves the classification effect of image aesthetics.

The rest of this paper is organized, as follows: The Section 2 reviews the related work. The proposed aesthetic assessment models are introduced in the Section 3. The Section 4 is comprehensive experiments. The Section 5 concludes this article.

2. Related Work

Image aesthetic has been studied for years, and a large number of aesthetic assessment methods have been reported in the literature. This section will focus on prior representative works.

2.1. Feature-Based Approach

In the early assessment of image aesthetics, the researchers focused on how to extract more effective manual features. For this reason, other fields or user-defined rules were introduced as aesthetic assessment criteria, including the aesthetic rules in the fields of photography and painting [16], rules that are based on human visual attention mechanisms [17], color harmony rules [18], content-aware retargeted rule bases on the structural similarity index [19], etc., and then the machine learning methods are used to perform high aesthetic perception and low aesthetic perception, such as two-category or aesthetic score prediction on the image. Datta [20] was the first researcher to propose extracting the low-level features and high-level features related to the aesthetics of the image, and SVM is used to classify these features. Obrador [21] explored the role of low-level composition features in image aesthetic classification, while using simpleness and visual balance to construct an image aesthetic classifier. Luo [16] proposed evaluating the aesthetic quality of photos based on the regional and global features of the content and the main body of the image that is most concerned by the human eye. The evaluation results depend on the image content and the extracted subject region. Starting from the characteristics, Wang [3] proposed a comprehensive feature set of a total of 86-dimensional features, including 16 new features and 70 proven effective features. Through machine learning, the classification accuracy rate reached 82.4%. Aesthetic assessment is subjective and difficult accurately model and quantify in engineering because the image aesthetics are ever-changing. Therefore, manual features often have an insufficient representation of aesthetic information, and it is difficult to fully express the aesthetics of images, but it is an approximate representation of aesthetic rules. Liu [22] proposed a novel semi-supervised deep active learning (SDAL) algorithm, which discovers how humans perceive semantically important regions from a large number of images partially assigned with contaminated tags. Zhang [23] proposed a Gated Peripheral-Foveal Convolutional Neural Network (GPF-CNN). It is a dedicated double-subnet neural network, i.e. a peripheral subnet and a foveal subnet. The former aims to mimic the functions of peripheral vision in order to encode the holistic information and provide the attended regions. The latter aims to extract fine-grained features in these key regions.

2.2. Methods Based on Color Distribution

Although some researchers have proposed different image aesthetic assessment models and frameworks, it is still a very challenging task to calculate the aesthetic assessment of images. Some researchers have proposed to use the color harmony model to evaluate images aesthetic to overcome problems of traditional manual features, according to the color distribution of images. Nishiyama [24] proposed a block-level color harmony model to extract color distribution features in a single color channel, and the recognition accuracy on the DPChallenge dataset is close to 66%. Lu [4] proposes using the EL-LDA (Extend Labelled-Latent Dirichlet Allocation) model to mine the color usage rules in the image. The accuracy of aesthetic assessment is higher than those models based on color harmony. The EL-LDA algorithm still focuses on the color distribution, which tries to explore the color combination rules of aesthetic images. Marchesotti [25] proposed classifying image aesthetic by using the general feature descriptors of images. The general feature descriptors are mainly used in the multi-classification field of object recognition, and the lack of explanatory and expansiveness in the field of aesthetic assessment. Guo [5] used the SIFT (Scale-invariant Feature TFransform) dense sampling feature descriptor [26] and LLC [27] to perform image semantic feature representation. Guo's method solved the problem of accurate aesthetic assessment when subject area extraction fails by combining manual features and semantic features to assess image aesthetic quality. It is proven that semantic features can improve the performance of image aesthetic assessment. However, the semantic feature dimension is too high, and the interpretation of aesthetic assessment is lacking. It is difficult to realize the real-time assessment of aesthetic images.

3. Method Formulation

This section might be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

3.1. Image Aesthetic Assessment Framework Based on Latent Semantic Features

It can improve the predictive ability of aesthetic assessment by the probability model—Latent Dirichlet Allocation Model [14]—mining the common features between different images from the image aesthetic database, as a generalized feature descriptor of aesthetic assessment. The generalized feature descriptor of the extracted image [25] helps to improve the subjective perception correlation between the aesthetic assessment model of computable images and human beings, given the different emphasis of human subjective evaluation. Mapping and coding representation of extracted image features can effectively characterize image information and be used for aesthetic assessment in machine learning [28]. This paper designs an aesthetic assessment framework, as shown in Figure 1, which consists of four modules, namely weighted feature extraction, feature mapping coding, semantic feature, latent semantic feature extraction, and classification identification, in order to capture the latent semantic features in image aesthetics and combine the re-encoding semantic features for aesthetic assessment.

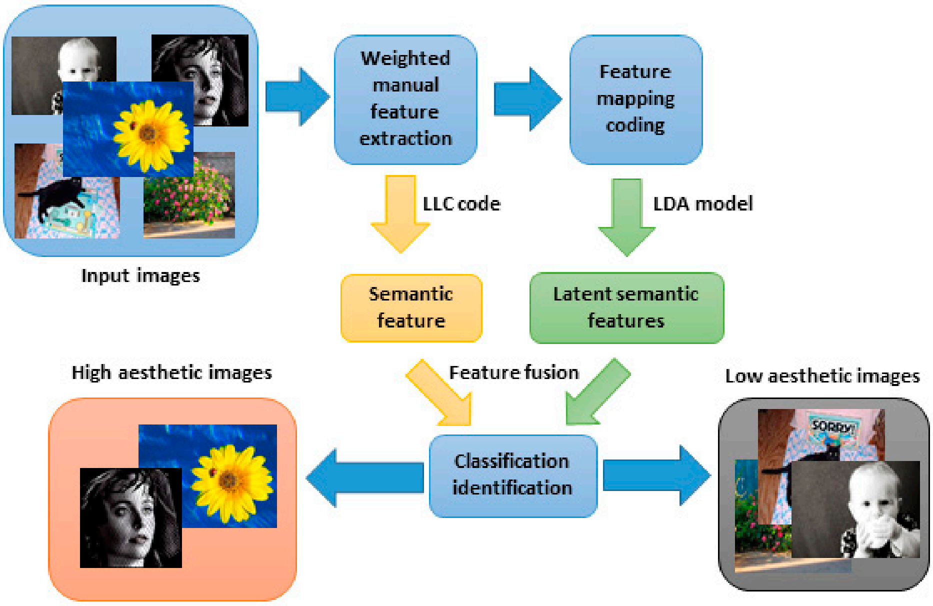


Figure 1. Flow chart of the proposed aesthetic assessment method.

The image segmentation method is the first step for preprocessing in this paper. There are serval methods for image segmentation, such as superpixel segmentation methods [29,30], watershed segmentation methods [31,32], and active contour methods [33,34]. Superpixel segmentation [35] is to divide adjacent pixels into irregular pixel blocks according to features, such as texture, color, and brightness. When compared to simple mesh sampling, superpixel partitioning is more practical. In this framework, first, the superpixel segmentation algorithm SLIC (Simple Linear Iterative Clustering) is used to segment the input image [35] into superpixel blocks. The POI is extracted according to the SIFT, SURF, or ORB algorithms, and the number of POI in the superpixel block is counted, and the point density processing is performed on the superpixel block. Seven kinds of manual features are extracted on the superpixel block, respectively, which include HSV(Hue, Saturation, Value) color histogram [36], color distance [3], Tamura texture features [37], LBP(Local Binary Pattern) features, Gabor features [37], GLCM (Gray Level Co-occurrence Matrix) features [37], and superpixel centroids descriptor to form visual vocabularies as aesthetic features of the image, according to the visual word bag model [28].

Subsequently, the semantic aesthetics features are acquired by encoding the extracted image aesthetic features using LLC code. The following operations are carried out in this paper in order to capture the aesthetic characteristic law of images: quantifying the extracted features, obtaining feature vocabulary documents through feature mapping coding, training LDA models, and mining potential topic vocabularies. Besides, latent semantic features are extracted by the similarity between feature documents and topic vocabularies.

Finally, during the process of image aesthetic classification assessment, this paper uses a support vector machine (SVM) [38] to train and test images in the AVA dataset [15], and the latent semantic features and different semantic features were combined for aesthetic score prediction.

3.2. Feature Extraction

This paper selects seven kinds of handcraft features of the image in order to fully reflect the characteristics of the image. Additionally, it extracts the aesthetics feature of the image using the local feature descriptor based on superpixel segmentation. The feature is extracted by simulating the process of image document representation to form a computable aesthetic assessment feature, and quantizing statistical encoding the image in order to obtain a feature set.

3.2.1. Superpixel Segmentation Based on Density Weighted POI

Image complexity is an important reference in image aesthetic assessment. The density of POI in images is taken as an important measure of complex local features, according to the principle of POI detection. Since the density of POI depends on the size of the partial block, in order to more reasonably represent the image features by block, this paper selects the SLIC algorithm to superpixel segmentation of the image and divides the image into irregular small blocks according to the similarity between pixels. The SIFT, SURF, and ORB algorithms are used to detect POI and count the number of POI in each block. According to Equation (1), the density of POI M_{δ_i} of the ith superpixel block δ_i of the image is calculated.

$$M_{\delta_i} = \frac{1}{N_{\delta_i}} \sum_{u \in (SIFT,SURF,ORB)} n_{u_i} \tag{1}$$

where N_{δ_i} indicates the area of the superpixel block and n_{u_i} indicates that the number of POI in a superpixel block, which is extracted by different methods, such as SIFT, SURF and ORB.

Set threshold T, when $M_{\delta_i} < T$, the weight of the superpixel block is 0, that is, there is no or fewer POI within superpixel block, indicating that the superpixel block is the background, and the extracted features have little effect on the aesthetic assessment. When $M_{\delta_i} > T$, it indicates that the background information in the block is sufficiently complex, which is largely the focus of attention of the human eye, directly affecting people's assessment of image aesthetics. Accordingly, it gives greater weight to the superpixel block, highlighting the importance of this area. Superpixel block weight W_{δ_i} is calculated as follows:

$$W_{\delta_i} = \begin{cases} 0 & if\ M_{\delta_i} < T \\ 1 + M_{\delta_i} & if\ M_{\delta_i} \geq T \end{cases} \tag{2}$$

Here we set T as mean of M_{δ_i} which is the optimal value according to a lot of experiments.

3.2.2. Superpixel Block Centroid Feature Descriptor

This paper uses SIFT, SURF, and ORB to extract feature descriptors, which can represent the semantic features of the image from different angles. Inspired by [5], we use the SIFT dense sampling feature descriptor to extract the semantic features of the image. In order to generalize the feature descriptor, the extracted feature descriptors are normalized.

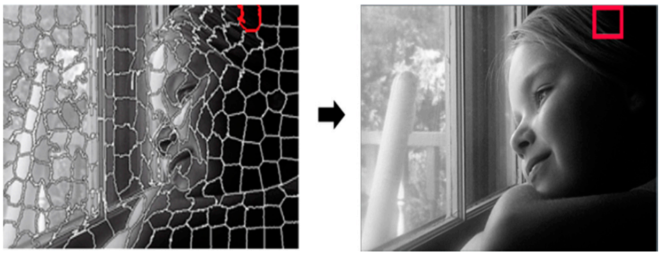
Using the centroid position of the superpixel block, the centroid feature descriptor of each superpixel block is extracted. According to the SIFT descriptor, the centroid of each superpixel block is extracted as a POI to extract the texture feature descriptor in the region of $U \times U$. The centroid calculation method for each superpixel block is as follows:

$$\begin{cases} x_{\delta_i} = \dfrac{\sum x \times I_{(x,y)} \in \delta_i}{\sum I_{(x,y)} \in \delta_i} \\ y_{\delta_i} = \dfrac{\sum y \times I_{(x,y)} \in \delta_i}{\sum I_{(x,y)} \in \delta_i} \end{cases} \tag{3}$$

where x_{δ_i} is the x axis coordinates of the ith superpixel block, y_{δ_i} is y axis coordinates of the ith superpixel block. $I_{(x,y)} \in \delta_i$ is the pixel inside δ_i in the image I and its value is 1 if it is inside of δ_i. Otherwise, its value is 0.

3.2.3. The Color Texture Feature Descriptor

The color feature and texture feature are extracted in superpixel blocks, respectively. The color texture feature descriptor is used to represent the aesthetic information of the image. The HSV color distance feature [3] and the color histogram feature [30] are used as color feature information of image aesthetics. In this paper, the color feature descriptors are extracted for each superpixel block and weighted according to the density of the detected POI while extracting the color features of the whole image. When calculating texture feature descriptors, the superpixel segmentation is performed on Tamura texture features, LBP texture features, Gabor wavelet transforms texture features, and GLCM. The centroid of the superpixel block is taken as the center of the circle, and R is taken as the radius, and the texture feature in the external rectangle of the circle is used to approximate the texture feature of the superpixel block in order to calculate the Gabor feature and the gray level co-occurrence matrix feature in the irregular superpixel block. The pixel block (see Figure 2b) can be determined by the superpixel block (see Figure 2a) to perform image segmentation, and the local Gabor feature and local gray level co-occurrence matrix are extracted.



(**a**) (**b**)

Figure 2. Gabor features and gray level co-occurrence matrix features extraction in superpixel blocks, (**a**) superpixel block diagram, and (**b**) centroid circle circumscribed rectangle of superpixel block.

3.3. Feature Mapping Coding

It is needed to map and encode the feature set extracted from each image in order to quantify the aesthetic features of an image. The mapping and encoding process is a further abstraction and general representation of the image features, enabling the machine to better understand the subjective and complex classification tasks of aesthetic assessment. Through feature mapping and encoding, the semantic features and latent semantic features are extracted, respectively.

3.3.1. Semantic Features

Semantic feature coding models, such as BOW [28], LLC [27], and so on, are widely used in pattern recognition models. This paper uses the LLC model to encode the feature descriptors. The LLC model improves the hard coding scheme in the bow model, which can better capture the difference information among different images, and the identification accuracy is higher than the BOW model.

The LLC coding is directly related to the dictionary codebook size of the feature vocabulary and it is obtained by the K-Means clustering method. In this paper, according to the feature descriptor D, the number of clusters is selected in order to improve the coding effect. K-means that the clustering method generally uses the Elbow Method to select the appropriate number of clusters. For convenience, the following method is used to select the number of clusters similar to the Elbow Method because the size of the feature descriptor set is different in magnitude.

$$Km = \begin{cases} 2048 & if\ N_D > 1 \times 10^7 \\ 1024 & if\ 1 \times 10^6 < N_D \leq 1 \times 10^7 \\ 512 & if\ 1 \times 10^5 < N_D \leq 1 \times 10^6 \\ 256 & if\ N_D \leq 1 \times 10^5 \end{cases} \tag{4}$$

where N_D represents the size of the feature descriptor D. All of the feature descriptors are encoded into corresponding semantic features by LLC coding.

3.3.2. Feature Mapping Coding

The feature descriptors are quantized into words by feature mapping coding to form a multi-semantic document of the image. Here, the feature descriptors of different dimensions are integrated to mine the latent semantic rules of image aesthetic features, capture common information in feature documents, and perform word mapping and encoding quantization on feature descriptors.

First, the extracted feature descriptors are uniquely numbered, and the feature numbers and feature code contents are distinguished by specific characters. When considering that the feature vector extraction methods are different, the representative meanings are different. A feature prefix is added in front of each feature word to distinguish the feature words in order to distinguish the feature types represented by the words. The feature prefix is composed of the feature number and specific characters.

Subsequently, the character representation of the quantization interval is defined, and the number of characters is consistent with the number of quantization intervals. For feature vectors, there are many cases of eigenvalues in each dimension, which is not conducive to statistics. When the accuracy is preserved as much as possible, the eigenvalues are quantized into M intervals. The intervals are respectively represented by specific symbols. The spatial mapping method is used for converting the feature vectors into words for subsequent word frequency statistics and similarity calculation. The specific mapping methods are as follows:

$$Cidx = ceil(\frac{x}{\max(\boldsymbol{X})} \times M) \tag{5}$$

where $Cidx$ represents a character index, $ceil(x)$ is to the round-up function, and M is the number of characters in the quantify interval. $\boldsymbol{X}$ represents a feature vector, and x represents the eigenvalue of the feature vector.

After the feature is quantized into a word, if the length of the feature word is too long, it is not conducive to statistical word frequency, mining semantic information, and the feature vector has consecutive zero values or consecutive identical feature values. The encoding method of feature words is further improved, and the feature vector is reduced in dimension, in order to solve these problems. This paper uses the OTSU algorithm to compress and encode long feature words.

Using the OTSU algorithm for reference, the eigenvalues of the feature vector are considered to contribute to the semantic coding representation component or the meaningless component. The OTSU algorithm determines the segmentation threshold in the feature vector. The eigenvalues larger than the threshold are encoded, and the less than the threshold is not encoded. The improved coding method is as follows:

$$Cidx = ceil(\frac{x - T_{OTSU}}{\max(\boldsymbol{X}) - T_{OTSU}} \times M) \tag{6}$$

Among them T_{OTSU} is the threshold of the feature vector determined by the OTSU algorithm.

Finally, the feature descriptor is transformed into a feature word document according to the dimension of the feature vector.

3.3.3. Latent Semantic Features

LDA is a document topic generation model. In the field of data mining and information retrieval, LDA is widely used to mine the relevance of document semantic topics and find the most relevant set of the first N topic words by means of probability from large to small in an unsupervised manner. In the LDA model framework, each document is considered to consist of a probability distribution of a set of subject words, and each subject word is composed of a polynomial probability distribution of words in the vocabulary set. In this paper, the image feature descriptors are regarded as feature words with aesthetic information, and the potential semantic relations of these words, that is, the co-occurrence of features, are found, and the aesthetic features are evaluated as being latent semantic features.

The LDA algorithm trains the co-occurrence rules and the distribution rules of potential subject words in the learning feature set, and then the latent semantic features in the multi-semantic feature document are extracted according to the distribution of the topic words. For the potential keywords mined, the similarity between the feature vocabulary of each image and the potential topic vocabulary is calculated, and the latent semantic feature vector of the image is obtained. The feature vector is determined by the potential topic word. In particular, all of the feature word conversions in the training and testing process follow a fixed set of mapping and encoding rules to ensure that the resulting latent semantic features are consistent.

After the feature set document is obtained, the potential semantic subject words are found by the LDA algorithm. The LDA algorithm has two key parameters: the number of subject words N and the number of topic content words K, that is, selected by the LDA algorithm N group subject vocabulary, each group has K feature words. The latent semantic features are obtained by statistically comparing the similarity between the feature words and topic vocabulary in the image semantic document. The similarity of words is measured by the Hamming distance, and a threshold is set t, less than t, then these words are considered to be similar, and the latent semantic feature extraction is as shown in Equation (7).

$$F_{LDA} = \frac{1}{sum(N_f)} \sum_{i=1}^{cout(N_f)} \sum_{j=1}^{N_{f_i}} V_{f_{ij}} \in T_f \ s.t. \ Hamdist(\forall T_f, V_f) \leq t \tag{7}$$

where $cout(N_f)$ is the number of kinds of feature vectors, $sum(N_f)$ represents the total number of characteristic words of the feature document corresponding to each image, and $Hamdist(\forall T_f, V_f)$ represents the Hamming distance between any subject vocabulary and feature words.

3.3.4. Assessment Model

Image aesthetic assessment is regarded as a machine learning problem. The process of machine learning is based on the process of how image features match the aesthetic image type and it then gives the probability of aesthetic classification prediction [39]. This paper uses the SVM classifier with kernel function RBF to establish an image aesthetic classification prediction model.

Find the optimal hyperparameter of SVM classifier by grid search method c (penalty factor, the tolerance for error) and γ (after selecting the RBF function as the kernel, a parameter that comes with the function). After selecting the optimal parameters and using them as parameters in cross-validation, this paper chooses to use ten times cross-validation to train the proposed model. During the training, the distribution of the training samples is randomly disturbed, so that the number of positive and negative samples in the sample set of each training is approximately 1:1, which avoids over-fitting during training in order to improve the generalization ability of the model.

The latent semantic features are features based on potential subject terms that are mined by LDA. According to unsupervised modelling of potential color harmony features based on GMM (Gaussian mixture model) [4], this paper carries out GMM training modelling on the latent semantic feature distribution, through the latent semantic features in the high and low aesthetic theme, the probability likelihood ratio of the word model is used to classify the image.

4. Experiment Database Settings

The software platform developed by the system is as follows: the operating system is Windows 7; using Visual Studio 2013 development tools for image feature extraction, cluster analysis, and parameter optimization, MATLAB R2014a is used for training classification experiments and ROC curve drawing. The development process used the OpenCV2.4.11 development kit and LIBSVM3.22 development kit.

4.1. Feature Extraction

The experimental database selected 6293 images from the AVA dataset as the training test set to find the optimal parameters of the experiment and the best classification accuracy in order to evaluate the effectiveness of the proposed method. Literature [14] provided a web download link to the AVA database, and a total of 255,348 images were collected. Each image is scored by more than 100 users. The score ranges from 1 to 10. The users who score the images come from different groups in different fields. They are not restricted by gender, age, and profession, ensuring the objectivity and importance of the ratings. The average score of each image is selected as the last image aesthetic score, and the score shift represents an increase in the beauty of the image, so it has an important guiding value in the field of image aesthetic evaluation. Figure 3 shows parts of the database pictures. The picture contains buildings, characters, animals, plants, landscapes, etc. The dataset is rich in content and representative.

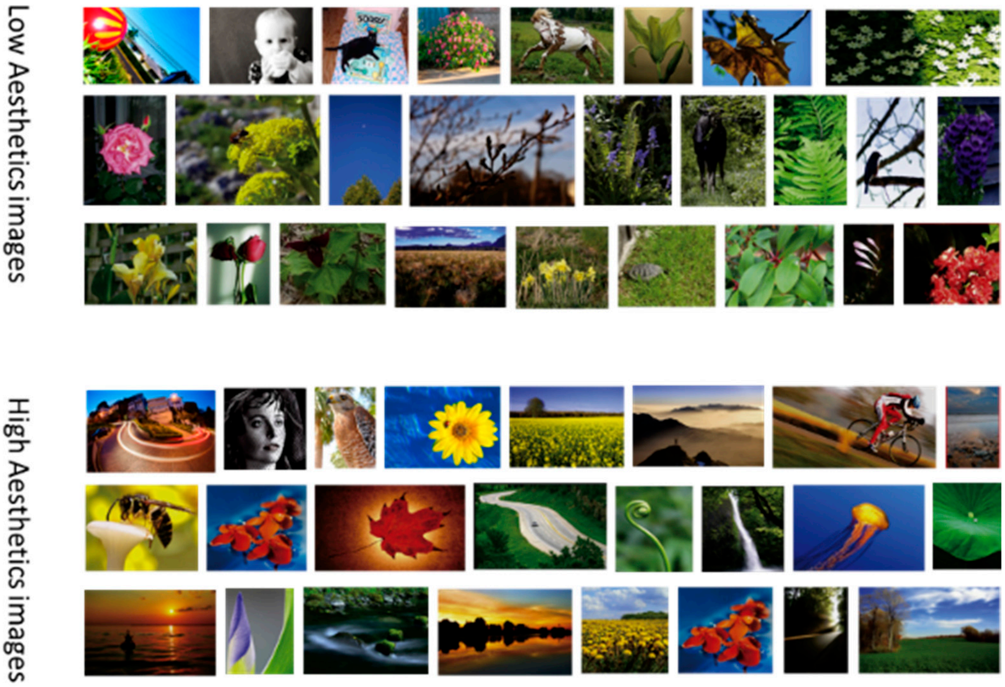


Figure 3. Images of high and low aesthetics quality.

We also conducted experiments on the Photo.net dataset in order to compare with other methods [39]. The Photo.net dataset is collected from www.photo.net. It consists of 20,278 images, but only 17,200 images have aesthetic label distribution. The distribution (counts) of aesthetics ratings is from one to seven.

This paper selects 10% data from both ends of the AVA dataset, and then takes the [4.5, 6.5] interval as the boundary, takes the datasets on both sides, selects 2000 high and low-quality images to train the model, and fully considers the selection process. Sample categories make the sample categories as uniform as possible. Then, a total of 2,492 images of 1,286 high-sensitivity images and 1,206 low-featured images were randomly selected from the remaining samples to test the model. In the end, from the 255,348 images that were collected, a total of 6,293 high and low aesthetic images were selected for experiments.

4.2. Experimental Steps

The image is subjected to superpixel segmentation and block processing by the SLIC algorithm to obtain a class label for each small block. Subsequently, the detection of POI is performed on the image, and the class label determines the superpixel block where POI is located. Afterwards, the superpixel block is weighted according to the distribution of the POI, and then the weighted local feature descriptor is extracted. K-means clustering is performed on the training set images to obtain a codebook dictionary set B of the feature set, and LLC encoding is performed according to the codebook dictionary B, and the semantic features of the image are obtained. Subsequently, according to the mapping and encoding rule presented in Section 3, the feature vector is converted into a word, and the latent topic word clustering is performed on the high and low aesthetic pictures by the LDA algorithm, and potential semantic features are extracted based on potential subject terms. This paper uses the ROC curve to evaluate the aesthetic assessment ability of latent semantic features and semantic features in order to evaluate the experimental performance.

We conducted experiments on different parameters in order to analyze the experimental performance. The parameters include the number of superpixel blocks, the K parameter size of local constraint coding, the codebook quantization size, and the potential subject size. We find the approximate optimal parameters through the grid search method, and the initial value of those parameters are listed in Table 1.

Table 1. Initial parameter setting.

Parameters	Feature Space (GMM)			Feature Space (SVM)		
	LDA	LLC (Descriptors)	SLIC (Local)	LDA	LLC (Descriptors)	SLIC (Local)
SLIC-count	300	300	300	300	300	300
Codebook-K	1024	1024	1024	1024	1024	1024
Topic-K	256	N/A	N/A	256	N/A	N/A
Topic size	64	N/A	N/A	64	N/A	N/A
Kernel-N	8	8	8	N/A	N/A	N/A

4.3. Parameter Analysis

Based on the initialization parameters, the accuracy is experimentally calculated. Table 2 shows the influence of the Gaussian kernel number parameter of the GMM model on the experimental results. From Table 2, the best Gaussian kernel number is 8 (marked in bold), and the highest accuracy is 69.07%. Through the training and classification test of latent semantic features, it can be found that the number of Gaussian kernels is highly correlated with the number of subject words and the size of the subject words. Subsequently, experiment with different superpixel partitioning block numbers, codebook size, several subject words, and number of words included in the subject words, select the optimal parameters, and use the SVM classifier for the training model, and select both RBF and Line kernel function. In the RBF kernel function, the optimal transcendental parameters are obtained through grid search c and γ. Table 3, Table 4, and Table 5, respectively, show the experimental results. For the SVM classifier, the super optimal parameters are used for different classification features with γ preferred. For the GMM algorithm, regularization parameters that correspond to different potential subject terms are also optimized.

Table 2. AVE(Average accuracy) in different Gaussian mixture models.

Kernel Number	LDA Feature Space (GMM)				
	128-64	256-64	256-128	512-64	512-128
4	0.6522	0.5074	0.5688	0.5323	0.6454
8	0.6807	0.6907	0.6338	0.5969	0.6805
16	0.6366	0.6009	0.5864	0.5227	0.6402
32	0.6221	0.6374	0.6783	0.5351	0.5949
64	0.6322	0.5022	0.6506	0.5748	0.5608

Table 3. AVE in different superpixel blocks.

SLIC-Count	SLIC Feature Space (SVM-RBF)					
	ColorDistance	**HSV**	**LBP**	**Gabor**	**Tamura**	**SLIC-CWD**
100	0.6338	0.6919	0.7525	0.7284	0.6450	0.6346
200	0.6478	0.7012	0.7834	0.7521	0.6623	0.6739
300	0.6590	0.6992	0.7838	0.7525	0.6839	0.6671
400	0.6450	0.7024	0.7942	0.7529	0.6903	0.6651

For the parameters of the superpixel block, Table 3 shows the experimental results. The trend is that the larger the number of superpixel blocks, the better, but it will slow down the processing speed of the algorithm. Through Table 3, it can be found that the accuracy rate has little change among the number of superpixel blocks in 200, 300, 400, and, finally, the number of blocks of 300 is selected for subsequent experimental parameter analysis. Table 4 shows the AVEs in different sizes of the codebook. Under different features, the codebook size corresponding to the optimal accuracy is oscillated in the range of 1024 and 1536. Although the best accuracy is obtained when the codebook size is 1536, when the codebook size is 1024, the optimal accuracy rate is the most, and the accuracy of the dense feature is similar to that obtained at 1536. This paper chooses 1024 as the optimal parameter for the codebook size experiments to deal with the problem of the time efficiency.

Table 4. AVE in different codebook sizes.

Codebook-K	Feature Space (SVM-RBF)						
	Dense	**SIFT**	**WSIFT**	**SURF**	**WSURF**	**ORB**	**WORB**
256	0.7545	0.6963	0.7076	0.6767	0.7084	0.5812	0.6502
512	0.7722	0.6959	0.7072	0.6747	0.7188	0.5913	0.6639
1024	0.7846	0.7208	0.7236	0.6927	0.7304	0.6053	0.6683
1536	0.7938	0.7196	0.7172	0.7092	0.7329	0.5820	0.6635
2048	0.7890	0.7202	0.7200	0.7020	0.7272	0.5880	0.6602

Table 5 shows the experimental results for the subject word size. When compared with the GMM algorithm and the SVM algorithm, the classification accuracy under the GMM algorithm is generally higher than the SVM algorithm, and when the number of subject words is 256 and the number of subject words is 64, the best result is 69.07%.

Table 5. AVE in different subject headings.

Topic-K	Topic Size 64		Topic Size 128	
	GMM	**SVM (Line)**	**GMM**	**SVM (Line)**
128	0.6807	0.6193	N/A	N/A
256	0.6907	0.6740	0.6783	0.6782
512	0.5969	0.6414	0.6805	0.6455

Under the optimal parameters of the SVM classifier, the results of the recognition accuracy after the single feature and all of the features are connected in series are shown in Table 6. Among them, the Dense-SIFT descriptor, the SLIC-LBP descriptor, the SLIC-Gabor descriptor, and the LDA latent semantic feature descriptor have a higher recognition rate. The accuracy of all features after concatenation is the highest, which is 84.28%, but the feature dimension very high. In this paper, the latent semantic features of LDA and other semantic features are concatenated and tested in order to find a better combination of features, verify the hypothesis proposed in this paper, and ensure that the feature dimension is not high. Table 7 shows the experimental results.

Table 6. Feature classification results.

Feature	Ave		
	Ave	**High**	**Low**
Dense-SIFT	0.7938	0.8033	0.7838
SIFT	0.7252	0.7411	0.7084
SURF	0.7333	0.7659	0.6984
SLIC-CWD	0.6558	0.6851	0.6247
HSV	0.6073	0.6159	0.5982
SLIC-HSV	0.7084	0.7053	0.7117
Color Distance	0.5945	0.6330	0.5534
SLIC-Color Distance	0.6843	0.6944	0.6736
Global Texture	0.7060	0.7084	0.7034
Tamura	0.5913	0.6244	0.5559
SLIC-Tamura	0.7100	0.7084	0.7117
LBP	0.6871	0.7356	0.6355
SLIC-LBP	0.7874	0.7939	0.7804
Gabor	0.6835	0.7084	0.6570
SLIC-Gabor	0.7898	0.8017	0.7771
GLCM	0.6033	0.6011	0.6056
SLIC-GLCM	0.5263	0.5583	0.4921
LDA	0.7930	0.7986	0.7871
All features	0.8428	0.8507	0.8343

Table 7. Combination of latent semantic features.

Feature	Dimension	Ave
SIFT	256 + 1024	0.8151
SURF	256 + 1024	0.8103
ORB	256 + 512	0.7926
SLIC-CWD	256 + 512	0.7882
HSV	256 + 64	0.8043
SLIC-HSV	256 + 512	0.8151
Color Distance	256 + 64	0.7906
SLIC-Color Distance	256 + 512	0.7942
Global Texture	256 + 64	0.8171
Tamura	256 + 64	0.7926
SLIC-Tamura	256 + 512	0.8010
LBP	256 + 64	0.8151
SLIC-LBP	256 + 512	0.8351
Gabor	256 + 64	0.8067
SLIC-Gabor	256 + 512	0.8379
GLCM	256 + 64	0.7874
SLIC-GLCM	256 + 512	0.7830

It can be found from Table 7 that all of the semantic features are concatenated with the latent semantic features, the recognition rate is greatly improved, the average accuracy is about 80%, and the highest recognition rate is 83.75%. After the combination of the SLIC-LBP feature descriptor and SLIC-Gabor feature description, the recognition accuracy can be equal to the recognition result of all features concatenated. After combining with the LBP feature descriptor and Gabor feature descriptor, the recognition accuracy is greatly improved. After the global texture features are combined, the average accuracy rate is 81.71%, but the feature dimension is only 320 dimensions. It is proven that the LDA feature that is proposed in this paper is effective, indicating that in the aesthetic assessment classification experiment, the combination of latent semantic features and semantic features can carry out image aesthetic assessment.

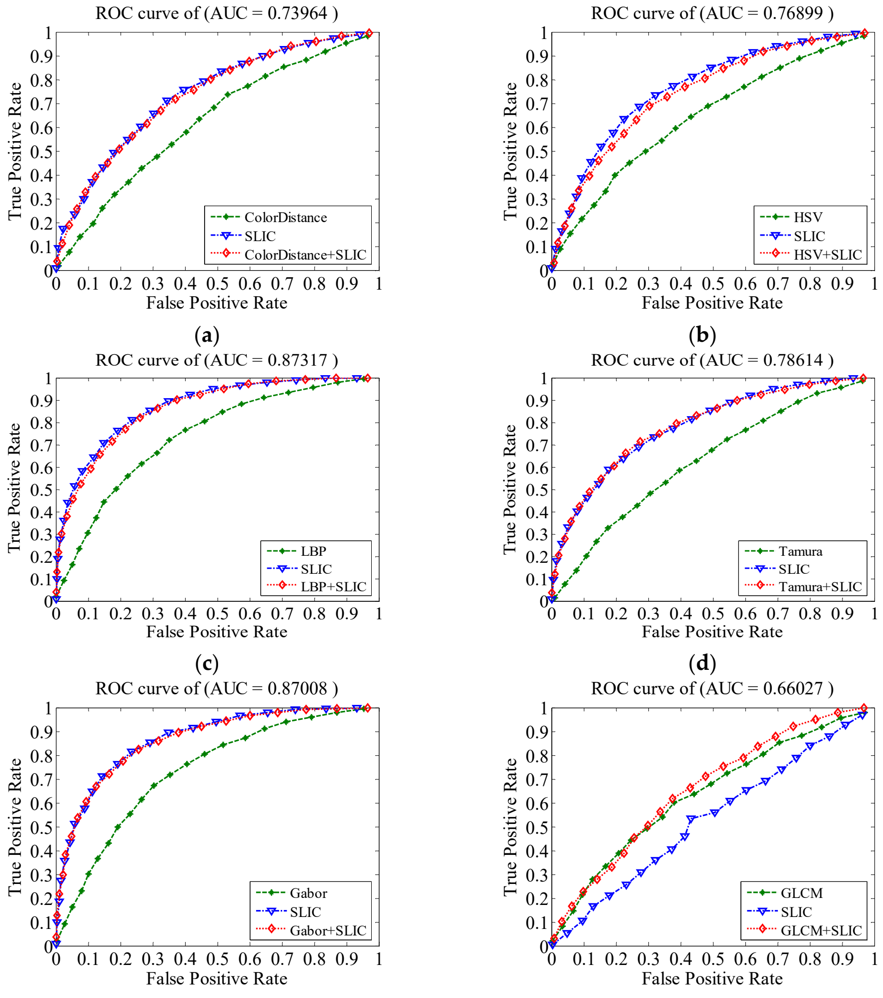
4.4. System Analysis and Comparison

In this section, the performance of the aesthetic assessment model that is based on latent semantic features proposed in this paper will be analyzed and compared with other aesthetic assessment models, including traditional heuristic-based methods and machine learning-based models.

4.4.1. System Output Analysis

The ROC (AUC) curve is used to prove the validity of the model through the visual analysis of the discriminative ability of the aesthetic assessment model proposed in this paper.

The ROC (AUC) curve that is shown in Figure 4 can prove that the proposed superpixel block weighting algorithm is effective and useful for improving the aesthetic classification assessment. Figure 4 shows a comparison of the experimental effects of six global features weighted by superpixel segmentation. The superpixel blocking weighting feature is represented by the blue line SLIC, and the red line represents the SLIC feature and global feature concatenating. It can be found that the weighted superpixel segmentation feature improves the classification performance of image aesthetics (except that the GLCM feature improvement is not significant), and the effect of the superpixel segmentation weighted feature effect and global feature stitching local feature is equivalent. The best classification performance is the LBP feature and the Gabor feature, and the AUC exceeds 0.87.



(**a**)

(**b**)

(**c**)

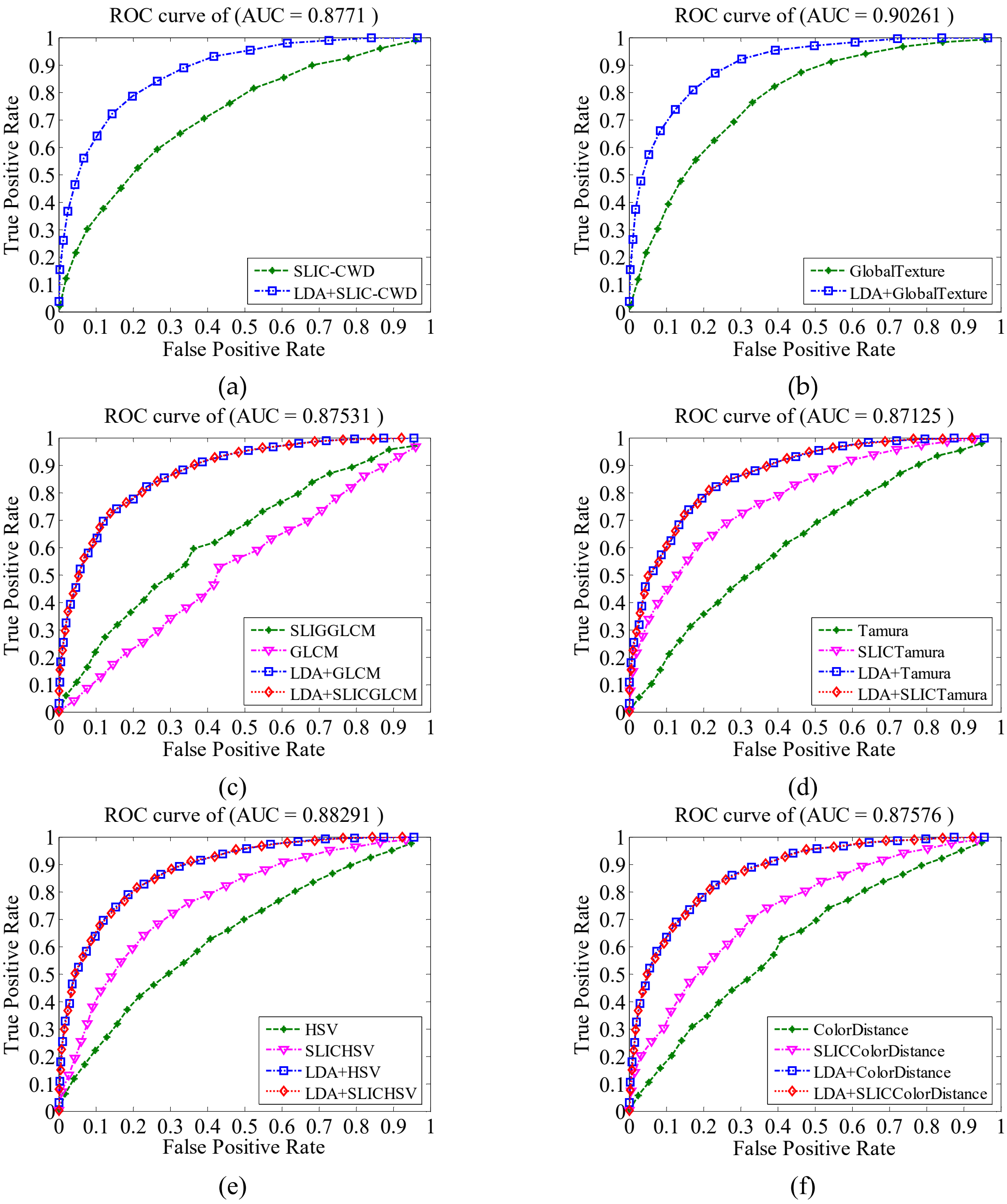
(**d**)

(**e**)

(**f**)

Figure 4. Comparison of the feature effects of global features after superpixel segmentation: (**a**) Color Distance feature, (**b**) HSV color histogram feature, (**c**) LBP feature, (**d**) Tamura texture features, (**e**) Gabor wavelet features, and (**f**) gray level co-occurrence matrix (GLCM) features.

Figure 5 shows the ROC (AUC) curve of the LDA latent semantic feature concatenated with the global feature and the corresponding superpixel block weighted local feature. Through the introduction of the latent semantic features of LDA, all of the features significantly improve the classification performance of image aesthetics. It is proved that the latent semantic features of LDA proposed in this paper can capture the feature co-occurrence information in image aesthetics, which can be used in the aesthetic assessment of images. Used with other semantic features to enhance aesthetic classification. The best classification performance is the combination of LDA features and global texture features (GlobalTexture); AUC reached 0.9026 and the other AUCs exceeded 0.87.



(a) (b) (c) (d) (e) (f)

Figure 5. ROC plot of LDA feature and semantic feature combination: (**a**) centroid descriptor feature (SLIC-CWD), (**b**) global texture feature, (**c**) gray level co-occurrence matrix (GLCM) feature, (**d**) Tamura texture features, (**e**) HSV color histogram features, and (**f**) color distance features.

4.4.2. Compared with Other Aesthetic Classification Models

The proposed algorithm is compared with state-of-art image aesthetic classification methods to prove the effectiveness of the proposed method. This paper chooses the classifier while using the LDA semantic feature and Gabor feature combination as the optimal model to compare with other methods, which includes Datta [20], Nishiyama [24], Marchesotti [25], Wang [3], Liu [22], and Zhang [23]. Among them, [22] and [23] are methods that are based on deep learning, and the others are handcraft feature methods. Table 8 shows the accuracy confusion matrix of the method. The average accuracy rate is 83.75%, the accuracy of high-sensitivity images is 85.15%, and the accuracy of low-quality images is 82.35%, which is the best in the binary classification.

Table 8. Method accuracy confusion matrix.

Acc	High	Low
High	0.8515	0.1485
Low	0.1765	0.8235

Table 9 shows a comparison between our method and state-of-the-art methods on the AVA dataset and Photo.net dataset, respectively. It is clear that our method has superior performance than conventional methods, like Datta [20], Nishiyama [24], and Marchesotti [25]. Besides, the recognition ratio of our method is 1.35% higher than that of Wang [3] et al., which illustrates the effectiveness of our method. Our method also has good performance compared with methods based on deep learning [22,23]. Especially when compared with [23], our method improves by over 10% on the Photo.net dataset and 1.94% on the AVA dataset. It also shows that our model can mine the co-occurrence relationship between image feature descriptions through latent semantic features, which is conducive to image aesthetic evaluation. Moreover, it can play a better role in the aesthetic evaluation of images by combining them with semantic features.

Table 9. Comparison experiment recognition ratio (%).

Methods	Photo.NET	AVA
Datta [20] (2006)	N/A	68.67
Nishiyama [24] (2011)	73.41	76.59
Marchesotti [25] (2011)	81.14	78.91
Wang W [3] (2016)	N/A	82.4
Liu Z [22] (2018)	87.56	83.09
Zhang [23] (2019)	77.19	81.81
Ours	87.78	83.75

5. Conclusions

In this paper, a model of potential semantic features for aesthetic classification is proposed. The LDA algorithm obtains the subject vocabulary of the feature document, and the similarity between the statistical feature word and topic vocabulary leads to the latent semantic feature. The latent semantic features are combined with other semantic features, and the best accuracy rate is 83.75% is obtained by SVM training classification. This paper provides a new solution to the image aesthetic assessment model, which encodes the aesthetic features into multiple semantic documents and mines the unspoken aesthetic information through the feature documents. The combination of invisible latent semantic features and visible semantic features greatly improves the image aesthetic classification effect. This paper proves that this direction is feasible through experiments. With the development of deep learning [22], for the complex classification task of image aesthetics, the deep semantics can be used to mine the latent semantic features for image aesthetic assessment. The method of extracting latent semantic features needs to be further improved and combined with other features to better evaluate the aesthetics of the image. It is still to be researched on how to more effectively extract the latent semantic features. The method of obtaining multi-semantic documents by word mapping coding method needs to be improved.

Author Contributions: The work presented in this paper represents a collaborative effort by all authors. Y.G. wrote the paper. R.B. and W.P. made a contribution to the latent semantic features. G.Y. analyzed the data and checked language. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of China under Grant number 61806071, the Natural Science Foundation of Hebei Province under Grant number F2019202381 and F2019202464, the Open Projects Program of National Laboratory of Pattern Recognition under Grant number 201900043, the Tianjin Sci-tech Development Strategy Research Planning Projects under Grant number 18ZLZXZF00660, and the Tianjin Technical Expert Project under Grant number 19JCTPJC55800 and 19JCTPJC57000.

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

References

1. Zha, Z.; Yang, L.; Mei, T. Visual query suggestion: Towards capturing user intent in internet image search. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2010**, *6*, 1–19. [CrossRef]
2. Obrador, P.; Oliveira, R.; Oliver, N. Supporting personal photo storytelling for social albums. In Proceedings of the 18th ACM International Conference on Multimedia, Firenze, Italy, 25–29 October 2010; pp. 561–570.
3. Wang, W.; Cai, D.; Wang, L. Synthesized computational aesthetic evaluation of photos. *Neurocomputing* **2016**, *172*, 244–252. [CrossRef]
4. Lu, P.; Peng, X.; Zhu, X. An EL-LDA based general color harmony model for photo aesthetics assessment. *Signal Process.* **2016**, *120*, 731–745. [CrossRef]
5. Guo, L.; Xiong, Y.; Huang, Q. Image esthetic assessment using both hand-crafting and semantic features. *Neurocomputing* **2014**, *143*, 14–26. [CrossRef]
6. Dong, Z.; Tian, X. Multi-level photo quality assessment with multi-view features. *Neurocomputing* **2015**, *168*, 308–319. [CrossRef]
7. Zhang, L.; Gao, Y.; Zimmermann, R. Fusion of multichannel local and global structural cues for photo aesthetics evaluation. *IEEE Trans. Image Process.* **2014**, *23*, 1419–1429. [CrossRef]
8. Aydın, T.O.; Smolic, A.; Gross, M. Automated aesthetic analysis of photographic images. *IEEE Trans. Vis. Comput. Graph.* **2015**, *21*, 31–42. [CrossRef]
9. Liu, C.; Chen, L.C.; Schroff, F. Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation. *arXiv* **2019**, arXiv:1901.02985.
10. Kim, C.; Shin, D.; Yang, C.N. Self-embedding fragile watermarking scheme to restoration of a tampered image using AMBTC. *Pers. Ubiquitous Comput.* **2018**, *22*, 11–22. [CrossRef]
11. Kligvasser, I.; Rott Shaham, T.; Michaeli, T. xUnit: Learning a spatial activation function for efficient image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Salt Lake City, UT, USA, 19–21 June 2018; pp. 2433–2442.
12. More, V.; Agrawal, P. Study on Aesthetic Analysis of Photographic Images Techniques to Produce High Dynamic Range Images. *Int. J. Comput. Appl.* **2017**, *159*, 34–38. [CrossRef]
13. Setchi, R.; Asikhia, O.K. Exploring User Experience with Image Schemas, Sentiments, and Semantics. *IEEE Trans Affect. Comput.* **2019**, *10*, 182–195. [CrossRef]
14. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
15. Murray, N.; Marchesotti, L.; Perronnin, F. AVA: A large-scale database for aesthetic visual analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Providence, RI, USA, 16–21 June 2012; pp. 2408–2415.
16. Luo, W.; Wang, X.; Tang, X. Content-based photo quality assessment. In Proceedings of the IEEE International Conference on Computer Vision(ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2206–2213.
17. Dhar, S.; Ordonez, V.; Berg, T.L. High level describable attributes for predicting aesthetics and interestingness. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Colorado Springs, CO, USA, 21–23 June 2011; pp. 1657–1664.
18. Moon, P.; Spencer, D.E. Geometric formulation of classical color harmony. *JOSA* **1944**, *34*, 46–59. [CrossRef]
19. Zhang, T.; Yu, M.; Guo, Y. Content-Aware Retargeted Image Quality Assessment. *Information* **2019**, *10*, 111. [CrossRef]
20. Datta, R.; Joshi, D.; Li, J. Studying aesthetics in photographic images using a computational approach. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 288–301.

21. Obrador, P.; Schmidt-Hackenberg, L.; Oliver, N. The role of image composition in image aesthetics. In Proceedings of the IEEE International Conference on Image Processing(ICIP), Hong Kong, China, 12–15 September 2010; pp. 3185–3188.
22. Liu, Z.; Wang, Z.; Yao, Y. Deep active learning with contaminated tags for image aesthetics assessment. *IEEE Trans. Image Process.* **2018**, 1. [CrossRef]
23. Zhang, X.; Gao, X.; Lu, W. A Gated Peripheral-Foveal Convolutional Neural Network for Unified Image Aesthetic Prediction. *IEEE Trans. Multimed.* **2019**, *21*, 2815–2826. [CrossRef]
24. Nishiyama, M.; Okabe, T.; Sato, I. Aesthetic quality classification of photographs based on color harmony. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Colorado Springs, CO, USA, 21–23 June 2011; pp. 33–40.
25. Marchesotti, L.; Perronnin, F.; Larlus, D. Assessing the aesthetic quality of photographs using generic image descriptors. In Proceedings of the IEEE International Conference on Computer Vision(ICCV), Barcelona, Spain, 6–13 November 2011; pp. 1784–1791.
26. Vedaldi, A.; Fulkerson, B. VLFeat: An open and portable library of computer vision algorithms. In Proceedings of the 18th ACM international conference on Multimedia, Firenze, Italy, 25–29 October 2010; pp. 1469–1472.
27. Wang, J.; Yang, J.; Yu, K. Locality-constrained linear coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 3360–3367.
28. Gandhi, A.; Alahari, K.; Jawahar, C.V. Decomposing bag of words histograms. In Proceedings of the IEEE Conference on Computer Vision (ICCV), Sydney, Australia, 3–6 December 2013; pp. 305–312.
29. Levinshtein, A.; Stere, A.; Kutulakos, K.N.; Fleet, D.J.; Dickinson, S.J.; Siddiqi, K. Turbopixels: Fast superpixels using geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *31*, 2290–2297. [CrossRef]
30. Stutz, D.; Hermans, A.; Leibe, B. Superpixels: An evaluation of the state-of-the-art. *Comput. Vis. Image Underst.* **2018**, *166*, 1–27. [CrossRef]
31. Gaetano, R.; Masi, G.; Poggi, G.; Verdoliva, L.; Scarpa, G. Marker-controlled watershed-based segmentation of multiresolution remote sensing images. *IEEE Trans. Geosci. Remote. Sens.* **2014**, *53*, 2987–3004. [CrossRef]
32. Cousty, J.; Bertrand, G.; Najman, L.; Couprie, M. Watershed cuts: Thinnings, shortest path forests, and topological watersheds. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *32*, 925–939. [CrossRef]
33. Ciecholewski, M.; Spodnik, J.H. Semi–automatic corpus callosum segmentation and 3d visualization using active contour methods. *Symmetry* **2018**, *10*, 589. [CrossRef]
34. Zhang, X.; Xiong, B.; Dong, G.; Kuang, G. Ship segmentation in SAR images by improved nonlocal active contour model. *Sensors* **2018**, *18*, 4220. [CrossRef] [PubMed]
35. Achanta, R.; Shaji, A.; Smith, K. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]
36. Nazir, A.; Ashraf, R.; Hamdani, T. Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor. In Proceedings of the IEEE International Conference on Computing, Mathematics and Engineering Technologies, Sukkur, Pakistan, 3–4 March 2018; pp. 1–6.
37. Deselaers, T.; Keysers, D.; Ney, H. Features for image retrieval: An experimental comparison. *Inf. Retr.* **2008**, *11*, 77–107. [CrossRef]
38. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans Intell. Syst. Technol.* **2011**, *2*, 1–27. [CrossRef]
39. Mentzer, F.; Agustsson, E.; Tschannen, M. Conditional probability models for deep image compression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Salt Lake City, UT, USA, 19–21 June 2018; pp. 4394–4402.