Personalized Image Aesthetics

Anonymous ICCV submission

Paper ID 244

Abstract

Automatic image aesthetics rating has received a growing interest with the recent breakthrough in deep learning. Although many studies exist for learning a generic or universal aesthetics model, investigation of aesthetics models incorporating individual user's preference is quite limited. In this work, we address this personalized aesthetics problem by showing individual's perception for image aesthetics is predictable and the features trained for generic aesthetics, aesthetics attributes and semantic content are all important to understand individual image aesthetics. To accommodate our study, we collect two distinct datasets, a large, Flickr image dataset annotated by Amazon Mechanical Turk, and a small dataset of real personal albums rated by owners. We also propose a personalized aesthetics model with a good generalizability that can be trained even with a very small set of annotated images from a user. The model is based on a residual-based model adaptation scheme which learns offsets to compensate generic aesthetics score for a user. Finally, we introduce an active learning algorithm for optimizing personalized aesthetics model prediction for real-world application scenarios. Experiments indicate that our approach outperforms previous recommendation-based active learning-based algorithms.

1. Introduction

Automatic assessment of image aesthetics is an important problem which has a variety of applications such as image search, photo editing and personal album curation [4, 22, 17]. It is a challenging task which requires a high-level, semantic understanding of a scene. Only recently, there has been a significant progress due to the advancement in deep learning that can learn such high-level information from data. Although many approaches have been proposed for generic or universal aesthetics prediction, research on automatic estimation of personalized image aesthetics is very limited. Since studies have demonstrated the subjective nature of individual aesthetics [36, 35, 32], thus directly applying the model trained for generic aesthetics to predict individual aesthetics is not accurate. For example, Figure 1 shows three images with aesthetics scores (between 1 to 5) from two different raters. Note that, for all three examples, the average score of 5 raters (which is regarded as the ground truth) is quite different from the score given by individual raters. These examples illustrate that a generic aesthetics model may not be sufficient to model individual user's aesthetics preference. Because different users may have very different preferences w.r.t. aesthetics attributes (composition, lighting, color) or contents of the image (portrait, landscape, pets). Therefore, understanding individual user's preference on aesthetics rating could significantly improve the accuracy for a personalized photo recommendation in an image search engine, and can be used to automatically curate personal albums.

(a) rater1: **3** rater2: **5** average for 5 raters: **3.6**
(b) rater1: **2** rater2: **5** average for 5 raters: **2.8**
(c) rater1: **1** rater2: **5** average for 5 raters: **3**

Figure 1: Examples illustrating personal aesthetics preferences. Each image is rated by 5 raters and scores of two are shown (15 raters for 3 images). As can be seen, individual users have very different aesthetics tastes on the same image and the average score of 5 raters, which is regarded as the ground truth generic aesthetics score, is far from representing individual users. Different users may pay attention to different attributes or contents in the images when they judge the aesthetics quality of an image.

Personalized image aesthetics refers to the problem of adapting a generic aesthetics model for individual user's preference, typically learned from a small set of aesthetics annotations provided by the user. In this study, we focus on how to solve the personalized image aesthetics problem. To this end, we need new datasets for training and evaluation. Existing aesthetics datasets are not appropriate for learning personalized image aesthetics as they either do not have rater identities or is of limited size[24, 10]. Therefore,

we collect two distinct datasets: (1) user-specific aesthetics dataset (UAD) which contains 40,000 Flickr images annotated by 210 Amazon Mechanical Turk (AMT)[1] turkers; (2) real personal album dataset (PAD) that contains 14 real user's photo albums with aesthetic scores given by the album owners. We use the UAD dataset for training generic aesthetics models, and use both datasets for studying personalized image aesthetics.

Given the new datasets, we first train a convolutional neural network (CNN) using UAD to predict generic aesthetic ratings. Because although aesthetics is subjective across individuals, people still share the common aesthetic tastes in the specific domain[1, 11]. So we use the generic model to reflect such common aesthetics tastes. Then we introduce a novel personalized aesthetics rating model to adapt aesthetics prediction to a user even with a very small number of annotations. The model is based on a offset-based model adaptation scheme which learns offsets to compensate generic aesthetics score for a user. Inspired by the studies[7, 18, 24] that aesthetics attributes and contents information can be adopted individually to achieve better classification results of generic aesthetics, we study neural network feature representations effective for personalized aesthetics learning. And find that features trained for generic aesthetics, aesthetics attributes, and semantic content are also all important for the personalized image aesthetics. Finally, we introduce an active learning algorithm for optimizing personalized aesthetics model prediction in real-world application scenarios such as interactive photo curation or image recommendation in search engines. We perform various experiments to demonstrate that our personalized aesthetics model outperforms existing recommendation and active learning-based methods.

Our main contributions are three-fold:

- Two new datasets for personalized image aesthetics research.
- A novel automatic personalized image aesthetics model with residual-based model adaptation.
- An active learning-based method for personal image aesthetics.

2. Related Work

Automatic aesthetics quality estimation Existing studies on image aesthetics prediction focus on extracting visual features from images and mapping the features to annotated aesthetics values [5, 7, 12, 18, 26, 9, 23]. Recent works based on CNNs have shown state-of-the-art performance on Aesthetics Visual Analysis dataset (AVA) [24, 10, 16, 17, 33]. In [16, 17], the authors show that using the patches from original images could improve the accuracy when classifying images as low aesthetics and high aesthetics. Mai *et al.* [20] propose an end-to-end model with adaptive spatial pooling to process original images directly without any cropping. Kong *et al.* [13] explore novel network architectures by incorporating aesthetic attributes and contents information of the images. However, these works mostly focus on learning generic aesthetics models on one or multiple datasets. By contrast, we study the personalized aspect of aesthetics and develop a personalized aesthetics model to learn the difference between generic aesthetics and individual aesthetics ratings.

Users-specific prediction Collaborative filtering has been a popular algorithm for learning personalized predictions. And matrix factorization is a common approach that serves as the basis for most collaborate filtering methods[14, 15]. A shortcoming of matrix factorization-based methods is that it has a strict prerequisite that the training set needs to involve users' assessments for various items. As a result, the rating for a novel item that is not contained in training set is not predictable. To overcome the limitations, several improvements have been introduced. For example, Rothe *et al.* [28] introduce the visual regularization to matrix factorization that regresses a new image query to a latent space, while Donovan *et al.* [27] use a novel feature-based collaborative filtering that transforms the features of new item to latent vectors. However, those approaches assume there are considerable overlaps among items rated by different users. This assumption is not valid for personalized aesthetics prediction, as each user would only rate their own personal photos.

Active learning for training set selection Active learning is an effective method to boost learning efficiency. It tries to select the most informative subset as training data from a pool of unlabeled samples. Samples with large uncertainties are chosen, whose ground-truth values are collected to update the models. However, most active learning methods deal with classification problems [31, 29, 30], and in this study, our model gives a continuous aesthetics score, which is a regression problem. Therefore, the studies on active learning for classification is not suitable for our study because evaluation of uncertainties for unlabeled samples is nontrivial in regression methods such as support vector regression, there is a risk of selecting non-informative samples which may increase the cost of labeling [34, 6]. There have been some studies to solve the active learning for regression problems, such as Burbidge *et al.* [2] propose an approach that can be used in regression methods by selecting unlabeled images with the maximal disagreement between the regressors, which are generated by ensemble learning algorithms. And Demir *et al.* [6] propose a multiple criteria active learning (MCAL) method that uses diversity of training samples and density of unlabeled samples. The active learning method introduced in our work differs from those existing works in that we define an ob-

[1] https://www.mturk.com

jective function to select unlabeled images by considering the diversity and the informativeness of the images that are directly related to personalized aesthetics.

3. Datasets

Existing datasets on image aesthetics either do not have rater identities[24, 10] or have limited number of annotated images[13], thus are not suitable for the personalized image aesthetics study. In this work, we collected two new datasets and will later release them promote research investigation.

User-specific aesthetics dataset (UAD). We collect 40,000 images from Flickr[2] and obtain the aesthetics score ratings through AMT. The aesthetics scores range from 1 to 5 representing the lowest to the highest aesthetics level. In order to have a better quality control, workers are asked to take a qualification test which contains 20 images with known aesthetics scores. Moreover, we randomly insert 3 images with ground truth scores TS into each task to regularly check workers' performance. Those images are chosen from the Aesthetics and Attributes Database (AADB) [13] and have scores either 1 or 5, as the aesthetics ratings on those obviously bad or good images are more consistent among users. Each worker gets a performance score IS from each ground truth image according to Equation 1,

$$\text{IS} = \begin{cases} 1 & \text{if } |WS - TS| < 2 \\ 0.5 & \text{if } |WS - TS| = 2 \\ 0 & \text{if } |WS - TS| > 2 \end{cases} \quad (1)$$

where WS is the score rated by the worker. We can then get an average performance score WP_i for each user by averaging the performance scores over all the images he or she annotated. Five workers with low performance scores as they do random labeling are excluded from the dataset to make the annotations more reliable.

Each image is rated by 5 different workers, with their anonymized identities recorded. Besides the individual ratings, we also get a weighted average aesthetic rating for each image IM_j, $IM_j = \frac{\sum_{i=1}^{5} WP_i \times WS_i}{\sum_{i=1}^{5} WP_i}$, which can be used to train generic aesthetics model.

In total, 210 workers participated the annotation of UAD. The average performance score for all the workers is 2.55±0.35 which demonstrates there common aesthetics tastes shared by different people [1], especially for those images with high/low aesthetics quality. To see the score variance for each image, we calculate the weighted standard deviation for each image, $\sqrt{\frac{\sum_{i=1}^{N} WP_i \times (WS_i - IM_j)}{\frac{N-1}{N} \sum_{i=1}^{N} WP_i}}$, where N is the number of workers. The scores of average weighted standard deviation for images in different score ranges are shown in Table 1. We can see the scores for images with relative high/low aesthetics (in the range of (1, 2) and (3, 4)) are lower than the images in the range of (2, 3) and (3, 4), which shows the uniqueness existing for individual perception for aesthetics [11].

[1, 5]	(1, 2)	(2, 3)	(3, 4)	(4, 5)
0.79±0.32	0.71±0.21	0.82±0.30	0.87±0.30	0.71±0.23

Table 1: Average weighted standard deviation for images in different score range.

In order to use UAD to validate the personalized aesthetics model, we select 37 workers and the 4,737 images they rated as our testing users and images. The number of images they labeled ranges from 105 to 171 (avg. = 137). The remaining 173 workers and their labeled images are used for training generic aesthetics models. With our data splitting, the training set does not have any images labeled by the testing workers, and vice versa, so that it can simulate the real use case where users only provide ratings on their own photos, and the algorithm cannot access those photos and ratings beforehand.

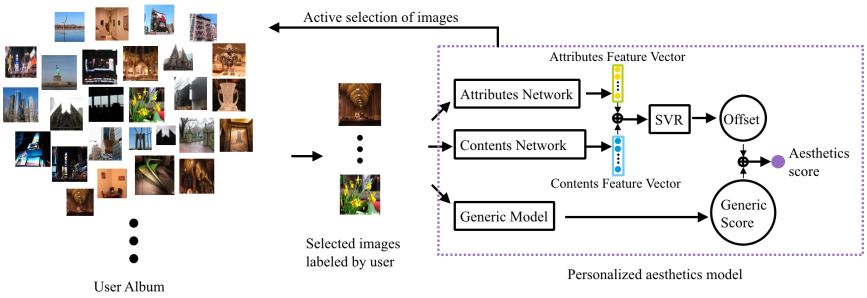
Compared with existing aesthetic datasets, UAD is the most comprehensive dataset with users' identities. AVA[24] and CUHKPQ [12, 19] only provide average image aesthetics scores. AADB[13] contains rater identities, but does not provide clear split of training and testing users. Moreover, UAD is four times larger than AADB, which we believe is not only more suitable for personalized image aesthetics, but can also be used to train a better generic aesthetics model.

Personal album dataset (PAD). In UAD, the ratings are not from photo owners. Therefore, to further evaluate personalized aesthetics in real-word applications, we collect another dataset composed of 14 personal albums and aesthetic ratings provided by the owners of the albums. The number of images within the albums ranges from 197 to 222 while the average number is 205. Similar to UAD, the aesthetics scores range from 1 to 5. To the best of our knowledge, this is the first aesthetics dataset with real users' ratings on their own photos available. We will release the two datasets upon acceptance of the paper. More details and example images regarding the datasets are shown in the supplementary materials.

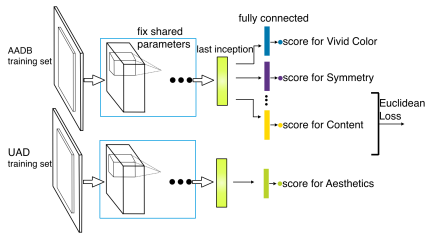
4. Approach

In this section, we propose our method to learn personalized image aesthetics using a small number of annotated images from each user. The pipeline is shown in Figure 2a. It consists of two main components: a personalized aesthetics model that estimates users' preference given a generic aesthetics model, and an active learning approach to select informative training images for more effective personalized aesthetics estimation in real applications.

[2]https://www.flickr.com



(a) Overview of active personalized aesthetics learning.



(b) A Siamese aesthetics attributes network for jointly training aesthetics and aesthetics attributes.

Figure 2: (a) We actively select the most informative images from user's own album and ask the user to label. Then the deep features of the training images are extracted to train a personalized aesthetics model. (b) The images with attributes scores from AADB and aesthetics scores from UAD are the ground truth input into the two streams. In each stream, the architecture is the same as Inception-BN except for the last inception. The intermediate inceptions are left out here for simplicity. The network calculates ten attributes scores and the aesthetics score.

4.1. Personalized aesthetics model (PAM)

In personalized image aesthetics, it is expected that a user can only provide a relatively small number of annotations on their personal photos. Therefore it is difficult to learn a model from scratch for each user to directly predict aesthetics ratings. Traditional recommendation approaches such as collaborate filtering may not be very effective as well, since there are very few overlapping images shared among users, and as a result the matrix formed by users and images are too sparse to learn strong latent vectors[27].

To solve the problem, we need to learn not only the common aesthetics taste that shared across individuals[1] but also the unique aesthetics perception of each individual [35, 32] . Following this idea, we propose to leverage the generic aesthetics model, and focus on the learning of the difference between individual aesthetics preferences and the generic aesthetics taste, which is easier to model with limited number of training data. In particular, the PAM includes two part. Firstly, we train a neural network to predict generic aesthetic scores with a large amount of training data. Secondly, when modeling the preference of a particular user, we run the generic model on the small set of images with provided annotations, and obtain the offsets between the user's ratings and the generic scores. Such score offsets reflect the user's preference on certain content or styles over common tastes. Hence we train an online ν-Support Vector Regression (SVR) model[3] to predict the offsets on the remaining images to adjust the aesthetic scores.

Generic aesthetics prediction. Previous methods[13, 16, 17] have demonstrated the capability of deep neural networks on modeling generic image aesthetics. Following these works, we train a trimmed Inception-BN to predict genetic aesthetic scores. It has the same architecture as in [8] except that we trimmed the number of neurons in the last inception (5b), which we found makes the training more efficient and yields better accuracies. The details of the last inception are illustrated in Table 2. Besides the trimmed inception, a dropout layer with 50% ratio is added after the last pooling layer. In order to give continuous aesthetics values, we use Euclidean loss to train the network, i.e.,

output size	$\#1 \times 1$	**double**$\#3 \times 3$ **reduce**
$7 \times 7 \times 176$	88	192
double $\#3 \times 3a$	**double**$\#3 \times 3b$	**Pool + proj**
224	56	max + 32

Table 2: The architecture of the last inception (5b) in our generic aesthetics model.

$$loss_{aes} = \frac{1}{2N}\sum_{i=1}^{N}\left\|y_i^{'} - y_i\right\|_2^2 \quad (2)$$

where y_i is the ground truth rating from AMT for the $i-th$ image, and $y_i^{'}$ is the corresponding predicted aesthetics score. Only one loss function is used at the end because we tried the multiple loss functions as in [8] but found the L2 loss-alone works the best for the UAD dataset.

Personalized aesthetics offsets prediction Given a generic aesthetics model and a personal album provided by a user, of which a small set has user's aesthetic ratings, we would like to adapt the generic model for this user and predict a personalized aesthetic offset over the generic aesthetic score for each image in the album. Previous studies have investigated the importance of incorporating attributes [7, 16, 22] and semantic content [18, 24] information into automatic assessment of generic aesthetics. Here we consider the attributes and content are also useful to model personalized perception. To model such preference, we use SVR with radial basis function kernel to predict personalized offset values as shown in Equation 3,

$$
\begin{aligned}
min \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C(\nu\epsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)) \\
s.t. \quad & (\mathbf{w}^T\phi(\mathbf{x}_i) + b) - y_i \leq \epsilon + \xi_i, \\
& y_i - (\mathbf{w}^T\phi(\mathbf{x}_i) + b) \leq \epsilon + \xi_i^*, \\
& \xi_i, \xi_i^* \geq 0, i = 1, ...l, \epsilon \geq 0.
\end{aligned} \tag{3}
$$

where $\mathbf{x}_i$ is the concatenation of aesthetic attribute features and content features, y_i is the target offset value, C is the regularization parameter, and ν ($0 < \nu \leq 1$) controls the proportion of the number of support vectors with respect to the number of total training images. We discuss the details of feature vector $\mathbf{x}_i$ in the following.

Specifically, AADB[13] has introduced ten attributes selected by professional photographers that related to image aesthetics and styles, including: *interesting content (IC)*, *object emphasis (OE)*, *good lighting (GL)*, *color harmony (CH)*, *vivid color (VC)*, *shallow depth of field (DoF)*, *rule of thirds (RoT)*, *balancing element (BE)*, *repetition (RE)* and *symmetry (SY)*. For the purpose of taking advantage of attributes features from images, we use the scores of those ten attributes ($att \in R^{10}$) as part of the input to the SVR model. In order to predict the attribute scores with limited number of training images provided by AADB, we leverage our trained generic aesthetics network, and propose a new Siamese attributes network that is jointly trained with attributes images from AADB and aesthetics images from UAD, as shown in Figure 2b. To learn the ten attributes and aesthetics scores, we use the joint loss function for the attributes network, as given by Equation 4,

$$loss = loss_{aes} + \sum_{i=1}^{N} w_i \times loss_{att_i} \tag{4}$$

where $loss_{att_i}$ is the Euclidean loss for the i_{th} attribute and w_i controls the relative importance of the attribute. The parameters of the attributes network are the same as the generic aesthetics network except for the last inception and the fully connected layers. Since the size AADB is four times smaller than UAD and we consider the image aesthetics and attributes should share similar low level features, so we only fine-tune the last inception of the Siamese attributes network from the generic aesthetics model while keeping parameters from other inceptions fixed.

As for the content features, we use the off-the-shelf image classification network[8] to extract semantic features (avg pool) from each image. Since the avg pool features are too long as the input to the offsets model, especially with a very limited number of training data, we utilize unsupervised k-means to cluster the images from the UAD training set into k semantic content groups using the semantic features and treat each cluster as a class. We then add a k-class soft-max layer on top of the network and fine-tune the layer with cross-entropy loss. Therefore, given an image, the output of the network would be a k-dimensional vector representing the probabilities belonging to each cluster. We use this k-dimensional vector as our content feature vector ($cont \in R^{10}$). Similar to [13], the value of k is set as 10. After concatenating the aesthetic attributes features and the content feature, we obtain the input feature vector to the offsets model $\mathbf{x} = [attr, cont]^T$, which is 20-dimensional. Experiments show the concatenation of attributes and content features achieve better results than using them alone.

Algorithm 1: Active-PAM

Input: Unlabeled photo set $N = \{p_1, p_2, ..., p_i, ..., p_N\}$, in which each photo is associated with image aesthetics feature aes_i, image aesthetics score $aescore_i$, the visual vector $v_i = [w_{aes} \times aes_i, aescore_i]$, $w_{aes} = 0.001$

Initialize the labeled photo set: $R = \emptyset$;

Randomly select a subset of ten images S from N, $N = N \setminus S$, $R = R \cup S$;

User rated aesthetics score t_i for $image_i$ in R, the number of images in R is indicated as $\#R$;

while *not stop labeling* **do**

  Iteratively train the regressor for offset using photos in R and get corresponding prediction set $\{o_i\}$;

  Calculate the weight for each training image

$$w_i = (1 - \frac{|o_i - t_i|}{\sum_{i=1}^{\#R}(|o_i - t_i|)}), p_i \in R \tag{5}$$

  Find p_q that

$$\max_q \sum_{j=1}^{\#R} w_j \times dist(v_q, v_j), p_q \in N, p_j \in R \tag{6}$$

  User label p_q with score t_q;

  $R = R \cup \{p_q\}$ and $N \setminus p_q$;

4.2. Active personalized image aesthetics learning (Active-PAM)

In real-world application scenarios such as interactive photo curation, users can continuously provide feedbacks regarding their aesthetics preference during the process[36, 35]. Instead of waiting for users to provide feedbacks on arbitrary images, we can pro-actively analyze the images in the album, select informative ones and ask the users to provide aesthetics ratings on those images. In this way, the PAM can be more efficiently learned with fewer training samples. To this end, we propose the Active-PAM to select/suggest training images for personalized learning.

When choosing the training images for model updating, we consider the following two criteria: 1) the images should cover diverse aesthetic styles, so that we can learn user's preference on different types of images; 2) the images with large offsets between user's ratings and the generic aesthetics scores are more informative. It indicates that the user's preference is different from common tastes on those images, and we would like to select similar images in the next round.

With these two criteria in mind, we implement Active-PAM as follows: for each image p_i in the album, its aesthetics score predicted by the generic aesthetics network is denoted by $aescore_i$, while the 176-dimensional output from the second to last layer in the network is denoted by aes_i. The aesthetic feature capturing the aesthetic styles of the image can then be represented as $v_i = [w_{aes} \times aes_i, aescore_i]$, where w_{aes} is a balancing factor, and is set to 0.001 based on the dimensions of the two outputs. Given the aesthetic features, we can measure the aesthetic dissimilarity between any two images p_i and p_j using the Euclidean distance $dist(v_i, v_j)$.

Assuming a set of images R that have already been selected and annotated by the user, for each remaining image p_i in the album, we can calculate the sum of distances between p_i and any image p_j in R, $d_i = \sum_{j=1} dist(v_i, v_j)$, $p_j \in R$. Apparently, the image with the largest d_i is the least similar one to the images in R in terms of aesthetic styles, and therefore satisfies the first criterion. In order to incorporate the second criterion at the same time, we would like to select the image that leans toward to the ones with larger aesthetic offsets in R. For each image p_j in R, we denote by o_j as its aesthetics score predicted by the current model, and denote its user-provided ground-truth score by t_j. We can then assign a weight w_j to each image using Equation 5. Obviously the weights are smaller when the offsets are larger. Hence we apply the weights when calculating the overall distance, resulting in Equation 6, where the distance/dissimilarity to the images with larger offsets are less considered. The images selected using Equation 6 are thus different from the ones that are already selected, and meanwhile could be more similar to the ones with larger aesthetic offsets, satisfying both criteria. The details of the active learning algorithm are described in Algorithm 1. The experiments show using the two criteria (diversity and informativeness) together achieves better results.

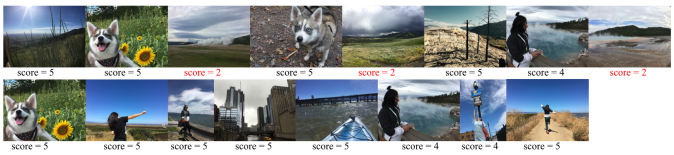
5. Experiments

In this section, we validate our approach by running it on three different datasets and comparing with several baseline methods. Figure 3 introduces visual examples to show how the personalized image aesthetics model works.

5.1. Generic aesthetics and attributes network

Implementation The aesthetics ratings given by AMT are scaled in the range of [0.2, 1]. The generic model is initialized from the Inception-BN [8] except the last trimmed inception and is fine-tuned on UAD. For the training process, images are warped to 256×256 and randomly cropped to 224×224 to feed into the network. The learning rate is initialized as 0.001 and periodically annealed by 0.96. The weight decay is 0.0002 and momentum is 0.9. In the testing stage, the predicted score is averaged on five crops (four corners and the center). The Siamese attributes network is fine-tuned from the generic model. The initial learning rate, weight decay and momentum are the same as the aesthetics generic aesthetics model. When jointly training attributes and aesthetics, we set w_i as 0.1 in Equation 4.

(a) Example from UAD



(b) Example from PAD

Figure 3: Two example results for personalized aesthetics prediction. The album in (a) comes from UAD and the album in (b) comes from PAD. The aesthetics scores under the images are given by user. In each example, top 8 images from different methods are shown here and scores smaller than 4 are marked with red color. First row: Aesthetics prediction from generic user-specific model; Second row: Aesthetics prediction from personalized aesthetics model.

	VC	GL	IC	SY	DoF
Baseline	0.5759	0.3770	0.4854	0.2283	0.5071
Our Results	**0.6938**	**0.4963**	**0.5641**	**0.2558**	**0.5476**
	OE	BE	CH	RE	RoT
Baseline	0.5728	0.2035	0.4808	0.3150	0.2174
Our Results	**0.6718**	**0.3104**	**0.5176**	**0.3749**	**0.2737**

Table 3: Attributes comparison of the baseline[13] and our results. Both calculated by correlation ρ. Jointly training attributes and aesthetics improves the attributes prediction.

Evaluation results The evaluation of generic model is measured by the Spearman's rank correlation (ρ) [25] between predicted aesthetics scores and the ground-truth values. The correlation is calculated as $\rho = 1 - 6\frac{\sum_{i=1}^{N}(r_i - r_i^{'})^2}{N^3 - N}$, where r_i is the rank of the i-th item when sorting the ground truth scores from high to low and $r_i^{'}$ is the rank for the estimated score. ρ ranges from -1 to 1 and higher value indicates better correlation between the predicted scores and the ground-truth scores. The correlation for the aesthetics generic model is 0.702. Table 3 shows the correlation on ten attributes. Interestingly, we found that the prediction of the attributes is significantly better than the one from[13], benefiting from larger training dataset in generic model.

5.2. Analysis of PAM

In this part, we analyze why aesthetics attributes and semantic content features are useful to learn personalized image aesthetics and compare PAM with a commonly used collaborate filtering approach, non-linear Feature-based Matrix Factorization (FPMF) introduced for individ-

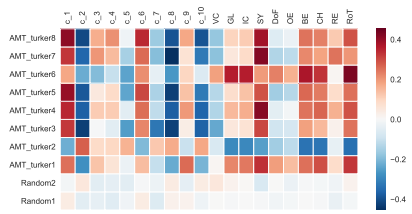


Figure 4: .

ual color aesthetics recommendation [27].

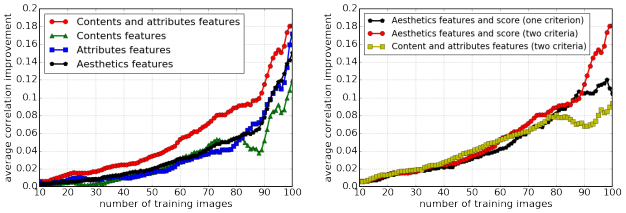
Does the personalized perception for aesthetics exist? In order to validate the effectiveness of PAM, we first demonstrate the existence of personalized image aesthetics. We generate m fake workers and randomly select k images from the training set as their annotated images. The score for each image is the ground-truth aesthetics score plus the random noise from a Gaussian distribution with mean as zero and standard deviation as σ. For all the images labeled by each worker, we calculate the summation of ranking correlation (ρ_{fi}) between each feature and offset values. For each worker j, we can get a score as $wc_j = \sum_{i=1}^{N} |\rho_{fi}|$, where N is 20 as shown in Section 4.1. Then we rank wc_j from low to high and give each worker a ranking score wc_{rankj}. So the higher the rank, the stronger the correlation between features and the offset values.

In the experiments, we set m as 10, test different values for k and σ and run the experiments for 50 times for each worker. We select 111 training workers from UAD who annotated more than 150 images (avg. = 1550). We average wc_{rank_j} of the ten fake workers and show the results in Figure 4. The average rank decreases with the increasing of images for the fake workers.

Image features for training PAM We compare the effectiveness of using different image features to regress the offset for the PAM. Figure 5a shows applying the concatenation of attributes and content features to train personalized aesthetics model is better than merely using contents, attributes or aesthetics features. The literatures have shown aesthetics attributes features and semantic content features can be used to improve generic aesthetics classification [7, 16, 18], our study further demonstrate the features can also be applied to the study of personalized aesthetics and the concatenation of the features represents better user's perception preference on aesthetics.

	SVM	FPMF	PAM
10 images	-0.352 $\pm$ 0.050	-0.001 $\pm$ 0.003	**0.006** $\pm$ **0.003**
100 images	-0.176 $\pm$ 0.064	0.010 $\pm$ 0.007	**0.039** $\pm$ **0.012**

Table 4: Comparison with SVM, non-linear FPMF[27] using different number of training images from each worker. The results are average correlation improvement.

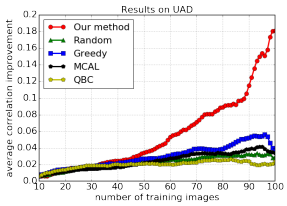


(a) Personalized aesthetics model trained with different features (with active learning).

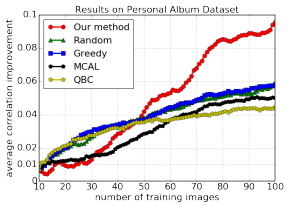
(b) Analysis of active learning using different criteria and features.

Figure 5: (a) We compare the correlation improvement when using different features to train personalized aesthetics model. The combination of contents and attributes features yields better results over other features. (b) We show the results of using one criterion (diversity) and two criteria (diversity and informativeness) to select samples. The visual vector is concatenation of aesthetics features and generic aesthetics score. We also show the results of using contents and attributes features as visual vector to calculate geometric distance (two criteria).

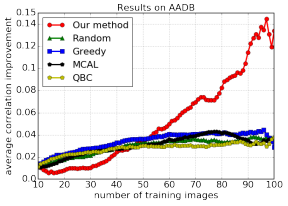
Quantitative results We apply the non-linear FPMF to learn personalized aesthetics in the experiments, which achieves better results than other matrix factorization approcahes[27]. The non-linear FPMF trains a one layer neural network to map the image features from training set into latent space, therefore it can predict user's preference on the images that are not contained in the training set[27]. The one layer neural network in non-linear FPMF includes 200 logistic units that can be trained by back-propagation and the dimensionality for the latent space is 15. We use the same features for non-linear FPMF as in PAM, i.e., the concatenation of attributes features and content features. We also train a support vector machine (SVM) as a baseline using the same features. We use the correlation improvement to evaluate those approaches. The correlation improvement is defined as the improvement of the ranking correlation ρ on the testing over the correlation achieved by the generic aesthetics model. The methods are compared on the testing users from UAD. Given a user, each time we randomly select k images as our training images to learn his/her personal preference, and test the model on the remaining images. For the non-linear FPMF, the training set also includes the 173 training users from UAD that used to train the generic aesthetics network. Due to the randomness of selecting training images, we run the experiments 50 times and report the average results as well as the standard deviation. The results with $k = 10$ and $k = 100$ are presented in Table 4. We can see that SVM and FPMF could not improve the ranking correlations with very small number of training images, and FPMF has marginal improvement even when using 100 training images. On contrary, the PAM works even with 10 training images, and has much more signifi-



(a) Testing results on UAD



(b) Testing results on PAD



(c) Testing results on AADB

Figure 6: We test Active-PAM on three different datasets. And we also compare Active-PAM with random selection, Greedy[37], MCAL[6] and QBC[2].

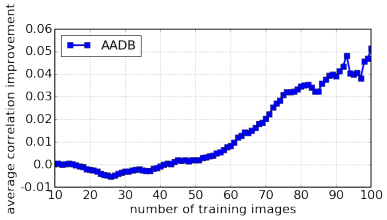


Figure 7: Average correlation improvement for AADB using the generic model from [13].

cant improvement with more training examples.

5.3. Analysis of Active-PAM

Visual vector for training set selection As described in Section 4.2, we use the aesthetic features to calculate the aesthetic dissimilarity between images and select images for active learning instead of using other features to encode image content and style information, (e.g, the content feature and attributes feature used to learn the personalized aesthetics model). We show the results using different features in active learning as in Figure 5b. The comparison demonstrates using generic aesthetics score and aesthetics features to select training samples is better than using attributes and contents features, as aesthetics features could better capture the diversity of images. We also show the results of using different criteria for the Active-PAM in Figure 5b, which demonstrate using two criteria (diversity and informativeness) is better than just considering one criteria.

Quantitative results In order to compare with other active learning approaches, we run the experiments on three different datasets, UAD, PAD and AADB. For the the evaluation on AADB, 28 workers who labeled 100 to 200 images are selected. We compare our method with other three active learning methods that focus on the study for regression. The Greedy[37] selects the unlabeled sample that has the minimum largest distance to the training samples at every iteration. The MCAL[6] chooses samples by clustering the unlabeled images and training images that are not support vectors. The Query by Committee (QBC) [2] generates a committee of models by using ensemble algorithms. In our experiment, we generate 5 committees using Bagging for QBC [21]. We did not compare with the study for classification because regression problem is different from the classification for active learning, such as the regressor gives continuous values which make the margin-based sampling strategy unsuitable[31, 6]. Results for 10 trials are shown in Figure 6. Our results achieve more significant correlation improvement on the datasets than others, especially when more training images become available.

5.4. Generalization study

Moreover, the Active-PAM can be applied to any generic aesthetic model to accommodate user preference. To validate the generalizability of Active-PAM, we use the network from [13] as the generic aesthetics model instead of our own trained network, and test the results of personalized image aesthetics. The reason for using the model developed in [13] instead of other generic models is that the model is trained on the dataset (AADB) that collected from natural images and with rater identity. So we can run the experiment on the same dataset by splitting images according to rater identity. We use 28 workers in AADB who label 100 to 200 images to test the performance. The result is shown in Figure 7, which indicates Active-PAM can be applied to different generic models for learning individual user's aesthetics model. The improvement is less significant than the one using our trained generic aesthetics network. It indicates that a better generic aesthetics model could help improve personalized image aesthetics more.

6. Conclusion

This work is the first attempt to study the subjectiveness of image aesthetics. We collect two new aesthetics datasets including rater identities and aesthetics scores that can be used to learn personalized image aesthetics. We train a CNN to estimate a generic aesthetics score and attributes scores. We then propose a personalized aesthetics model for accommodating individual aesthetics taste with limited annotated images. We also find that the attributes and contents are both important information for studying individual aesthetics preference. Furthermore, we introduce a new active learning method to interactively select training images and improve the training efficiency and performance of the personalized aesthetics model. One interesting future work is to investigate additional cues such as content redundancy, image quality or face recognition for improving user expe-

rience in real-world applications.

References

[1] P. M. Bronstad and R. Russell. Beauty is in the we of the beholder: Greater agreement on facial attractiveness among close relations. *Perception*, 36(11):1674–1681, 2007. 2, 3, 4

[2] R. Burbidge, J. J. Rowland, and R. D. King. Active learning for regression based on query by committee. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 209–218. Springer, 2007. 2, 8

[3] C.-C. Chang and C.-J. Lin. Training nu-support vector regression: theory and algorithms. *Neural Computation*, 14(8):1959–1978, 2002. 4

[4] B. Cheng, B. Ni, S. Yan, and Q. Tian. Learning to photograph. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 291–300. ACM, 2010. 1

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *European Conference on Computer Vision*, pages 288–301. Springer, 2006. 2

[6] B. Demir and L. Bruzzone. A multiple criteria active learning method for support vector regression. *Pattern recognition*, 47(7):2558–2567, 2014. 2, 8

[7] S. Dhar, V. Ordonez, and T. L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1657–1664. IEEE, 2011. 2, 4, 7

[8] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. 4, 5, 6

[9] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, and J. Luo. Aesthetics and emotions in images. *IEEE Signal Processing Magazine*, 28(5):94–115, 2011. 2

[10] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1733–1740, 2014. 1, 2, 3

[11] R. Kaplan and S. Kaplan. *The experience of nature: A psychological perspective*. CUP Archive, 1989. 2, 3

[12] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426. IEEE, 2006. 2, 3

[13] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes. Photo aesthetics ranking network with attributes and content adaptation. In *European Conference on Computer Vision*, pages 662–679. Springer, 2016. 2, 3, 4, 5, 6, 8

[14] Y. Koren, R. Bell, C. Volinsky, et al. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009. 2

[15] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001. 2

[16] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang. Rapid: rating pictorial aesthetics using deep learning. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 457–466. ACM, 2014. 2, 4, 7

[17] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 990–998, 2015. 1, 2, 4

[18] W. Luo, X. Wang, and X. Tang. Content-based photo quality assessment. In *2011 International Conference on Computer Vision*, pages 2206–2213. IEEE, 2011. 2, 4, 7

[19] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *European Conference on Computer Vision*, pages 386–399. Springer, 2008. 3

[20] L. Mai, H. Jin, and F. Liu. Composition-preserving deep photo aesthetics assessment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 497–506, 2016. 2

[21] N. A. H. Mamitsuka. Query learning strategies using boosting and bagging. In *Machine Learning: Proceedings of the Fifteenth International Conference (ICML'98)*, volume 1. Morgan Kaufmann Pub, 1998. 8

[22] L. Marchesotti, N. Murray, and F. Perronnin. Discovering beautiful attributes for aesthetic image analysis. *International Journal of Computer Vision*, 113(3):246–266, 2015. 1, 4

[23] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *2011 International Conference on Computer Vision*, pages 1784–1791. IEEE, 2011. 2

[24] N. Murray, L. Marchesotti, and F. Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2408–2415. IEEE, 2012. 1, 2, 3, 4

[25] J. L. Myers, A. Well, and R. F. Lorch. *Research design and statistical analysis*. Routledge, 2010. 6

[26] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato. Aesthetic quality classification of photographs based on color harmony. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 33–40. IEEE, 2011. 2

[27] P. O'Donovan, A. Agarwala, and A. Hertzmann. Collaborative filtering of color aesthetics. In *Proceedings of the Workshop on Computational Aesthetics*, pages 33–40. ACM, 2014. 2, 4, 7

[28] R. Rothe, R. Timofte, and L. Van Gool. Some like it hot-visual guidance for preference prediction. *arXiv preprint arXiv:1510.07867*, 2015. 2

[29] G. Schohn and D. Cohn. Less is more: Active learning with support vector machines. In *ICML*, pages 839–846. Citeseer, 2000. 2

[30] B. Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52(55-66):11, 2010. 2

[31] S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *Journal of machine learning research*, 2(Nov):45–66, 2001. 2, 8

[32] E. A. Vessel, J. Stahl, N. Maurer, A. Denker, and G. Starr. Personalized visual aesthetics. In *IS&T/SPIE Electronic Imaging*, pages 90140S–90140S. International Society for Optics and Photonics, 2014. 1, 4