# Personalized Image Aesthetics

Jian Ren[1] Xiaohui Shen[2] Zhe Lin[2] Radomír Měch[2] David J. Foran[1]
[1]Rutgers University [2]Adobe Research
jian.ren0905@rutgers.edu, foran@cinj.rutgers.edu {xshen, zlin, rmech}@adobe.com

## 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. We address this personalized aesthetics problem by showing that individual's aesthetic preferences exhibit strong correlations with content and aesthetic attributes, and hence the deviation of individual's perception from generic image aesthetics is predictable. To accommodate our study, we first collect two distinct datasets, a large image dataset from Flickr and annotated by Amazon Mechanical Turk, and a small dataset of real personal albums rated by owners. We then propose a new approach to personalized aesthetics learning that can be trained even with a small set of annotated images from a user. The approach is based on a residual-based model adaptation scheme which learns an offset to compensate for the generic aesthetics score. Finally, we introduce an active learning algorithm to optimize personalized aesthetics prediction for real-world application scenarios. Experiments demonstrate that our approach can effectively learn personalized aesthetics preferences, and outperforms existing methods on quantitative comparisons.*

## 1. Introduction

Automatic assessment of image aesthetics is an important problem that has a variety of applications such as image search, creative recommendation, photo ranking and personal album curation, etc. [16, 20]. It is a challenging task that requires high-level understanding of photographic attributes and semantics in an image. Only recently, there has been a significant progress due to the advancement in deep learning that can learn such high-level information effectively from data [15, 16, 30]. Although many successful deep learning-based approaches have been proposed for learning generic aesthetics classifiers, efforts on learning user-specific aesthetic models are quite limited.

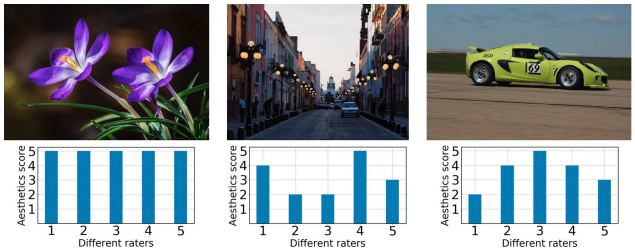



Figure 1: Examples illustrating personal aesthetic preferences. Each image is rated by 5 different users. As can be seen, on the first example, ratings are consistent among the users while on the other two, different users manifest very different visual tastes. Individual users may have their unique preference with respect to visual attributes or contents of an image when they judge its aesthetic quality.

Image aesthetics rating is well known to be a highly subjective task as individual users have very different visual preferences. This has been observed in earlier rule-based photo ranking systems such as [29, 33, 34]. For example, Figure 1 shows example images and their aesthetic scores (1 to 5) rated by five different users following careful instructions. As can be seen, ratings for each image could vary significantly among the users. This is expected as different users may have very different opinions with respect to photographic attributes (composition, lighting, color) or semantic contents of the image (portrait, landscape, pets). Therefore, learning individual user's visual preference is a crucial next step in image aesthetics research.

We refer to this problem as **personalized image aesthetics** and aim to address it by adapting a generic aesthetics model for individual user's preference. However, this is a challenging task as we typically need to learn such preference from a very limited set of annotated examples from each user. For example, in photo organization softwares, it is desired to minimize user's annotation effort. An effective strategy in such applications is to leverage active learning by automatically selecting representative images and suggest them to users for rating.

For investigating this problem with a learning-based approach, a labeled dataset with rater's identities are needed for training and evaluation purposes. Existing aesthetics datasets are not appropriate as they either do not have rater identities [22] or are limited in size [9]. Therefore, we introduce two new datasets specifically tailored for this task: (1) Flickr Images with Aesthetics Annotation Dataset (FLICKR-AES) which contains 40,000 Flickr images labeled by 210 unique Amazon Mechanical Turk (AMT)[1] annotators; (2) Real Album Curation Dataset (REAL-CUR) that contains 14 real user's photo albums with aesthetic scores provided by the album owners.

To learn the average aesthetic preference of a typical user [1, 10], we first use the relatively large FLICKR-AES dataset to train a powerful, generic aesthetics prediction model which performs competitively to the state of the art. Then, we present a novel approach for personalizing image aesthetics by adapting the generic model to individual users. To cope with a limited number of annotated examples from a user, we adopt a residual-based model adaptation scheme which is designed to learn a scalar offset to the generic aesthetic score for each user.

Inspired by the studies [5, 17, 22] which leverage both aesthetic attributes and content information to improve the performance of generic aesthetics rating, we study feature representations effective for personalized aesthetics learning, and observed that features trained for generic aesthetics prediction, aesthetics attributes classification, and semantic content classification are all important for learning personalized image aesthetics models. We further show in experiments that our personalized aesthetics model with this combined feature representation significantly outperforms an existing collaborative filtering-based method.

Finally, we introduce an active personalized aesthetics learning algorithm for real-world application scenarios such as interactive photo album curation. Results demonstrate that our method compares favorably to typical active learning-based methods in the previous literature.

Our main contributions are three-fold:

- We address the problem of personalized image aesthetics, and introduce two new datasets to facilitate research in this direction.
- We systematically analyze correlation between user ratings and image contents/attributes, and propose a novel approach for learning a personalized image aesthetics model with a residual-based model adaptation scheme.
- We propose an active personalized image aesthetics learning algorithm for real-world image search and curation applications.

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

## 2. Related Work

**Aesthetic quality estimation** Earlier studies on image aesthetics prediction mainly focus on extracting handcrafted visual features from images and mapping the features to annotated aesthetics labels by training classifiers or regressors [5, 11, 17, 21]. With the emergence of large-scale aesthetics analysis datasets such as AVA [22], a significant progress has been made on automatic aesthetics analysis by leveraging deep learning techniques [9, 15, 16, 31]. In [15, 16], the authors show that using the patches from original images could consistently improve the accuracy for aesthetics classification. Mai *et al.* [18] propose an end-to-end model with adaptive spatial pooling to process original images directly without any cropping. Kong *et al.* [12] explore novel network architectures by incorporating aesthetic attributes and contents information of the images. However, all these works focus on learning generic aesthetics models.

**Personalized prediction** Collaborative filtering has been a popular algorithm for recommendation and learning personalized preferences. Matrix factorization is a common approach that serves as the basis for most collaborate filtering methods [7, 13, 14]. Matrix factorization-based methods are strictly limited to existing items already rated by some users and cannot be used to predict/recommend novel items for users. To overcome the limitations, several improvements have been introduced. For example, Rothe *et al.* [25] introduce the visual regularization to matrix factorization that regresses a new image query to a latent space, while Donovan*et al.* [24] use a novel feature-based collaborative filtering that transforms the features of new item to latent vectors. Nevertheless, those approaches assume there are considerable overlaps among items rated by different users. In personalized image aesthetics, the sets of items rated by individual users may not necessarily be overlapping. For example, for photo curation, each user only rate their own personal images. Some earlier works on photo ranking [33, 34] incorporate user feedback in the ranking algorithms but it is done by adjusting feature weights in an ad-hoc way instead of learning from data.

**Active learning** Active learning is an effective method to boost learning efficiency by selecting the most informative subset as training data from a pool of unlabeled samples. Samples with large uncertainties are likely to be chosen, whose ground-truth values are collected to update the models. However, most active learning methods deal with classification problems [26, 27, 28], and in this study, our model aims to predict a continuous aesthetic score, which is formulated as a regression problem. Existing active classification approaches are not directly applicable to our problem because evaluation of uncertainties for unlabeled sam-

ples is nontrivial in regression methods such as support vector regression. Moreover, there is a risk of selecting non-informative samples which may increase the cost of labeling [4, 32]. There have been a few attempts to apply active learning for regression problems, such as Burbidge *et al.* [2] which select unlabeled images with the maximal disagreement between multiple regressors generated from ensemble learning algorithms. Demir *et al.* [4] 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 works in that we define an objective function to select unlabeled images by considering the diversity and the informativeness of the images that are directly related to personalized aesthetics.

## 3. Datasets

**FLICKR-AES** We download 40,000 photos with a creative commons license from Flickr[2] and collect their aesthetic ratings through AMT. The raw aesthetics scores range from 1 to 5 representing the lowest to the highest aesthetics level. Each image is rated by five different AMT workers and its ground truth aesthetics label is approximated to be the average of the five scores. In total, 210 AMT workers participated in the annotation of FLICKR-AES.

We split the dataset into training and testing sets. Specifically, we select 4,737 images labeled by 37 AMT workers to include in the testing set. The number of photos each testing worker labeled ranges from 105 to 171 (avg. = 137). All the remaining 35,263 images annotated by the rest 173 workers are included in the training set. We leverage the latter for training both generic and personalized aesthetics models. With this split, we verified that the testing set does not have any images labeled by the workers in the training set, and vice versa. This allows us to simulate real application scenarios where each user only provide ratings on his or her own photos, and the algorithm cannot access those photos and ratings beforehand.

Compared to the existing aesthetics dataset with rater identities [12], FLICKR-AES is a much larger and more comprehensive dataset which has a more diverse and balanced coverage of contents.

**REAL-CUR** In FLICKR-AES, aesthetics ratings are provided by AMT workers instead of actual owners of the photos in the dataset. For testing in the context of real-world personal photo ranking and curation applications, we collect another dataset composed of 14 personal albums (all from different people) and corresponding aesthetic ratings provided by the owners of the albums. The number of photos in each album ranges from 197 to 222 while the average is 205. As we only have one user rating for each photo, we instructed each user to go through their album multiple times to make the ratings consistent. To the best of our knowledge, this is the first aesthetics analysis dataset with real users' ratings on their own photos.

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

## 4. Analysis of User Preferences

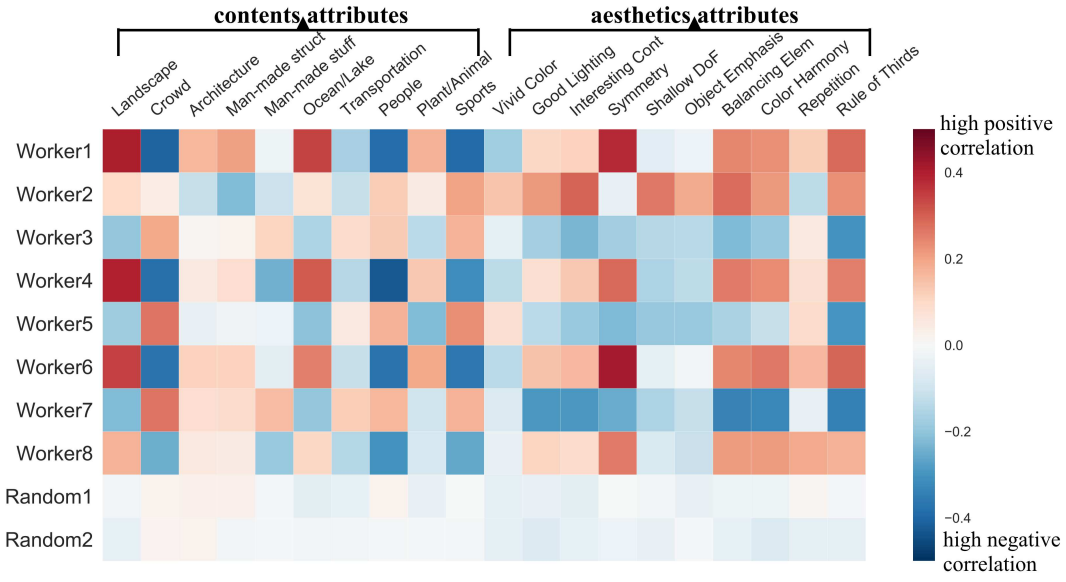
How significant are individual user's preferences relative to generic aesthetic perception? How are those preferences related to image attributes? In order to answer these questions, we perform the following correlation analysis between individual user's ratings and various image attributes using FLICKR-AES.

There are numerous image properties or attributes which could affect a user's aesthetics rating. Among them, we choose content attributes (semantic categories) and aesthetics attributes (e.g. rule-of-third, symmetry) as they are the most representative cues explored for aesthetics analysis and are proved effective for predicting aesthetic quality of an image[5, 17]. We show details on how to extract those attributes in Section 5.1.

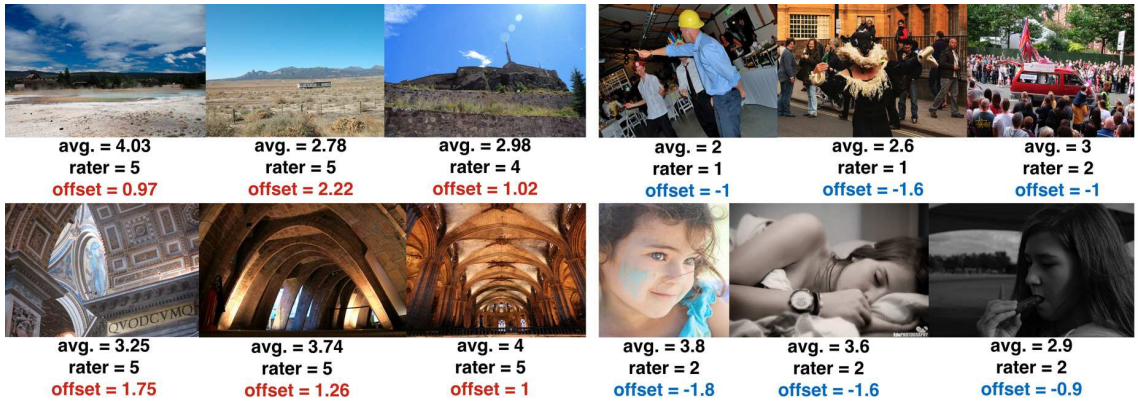
We select 111 AMT workers from the training set who have labeled at least 172 images (avg. = 1,550), and treat them as individual users. For each user, to measure the correlation of the preference and the content/aesthetics attributes, we use Spearman's rank correlation ($\rho$) [23] which is calculated 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 scores given by the first metric in descending order and $r_i^{'}$ is the rank for the scores given by the second metric. $\rho$ ranges from -1 to 1 and a higher absolute value indicates stronger correlation between the first metric and the second metric.

Directly measuring the correlation between user's absolute aesthetics scores and image attributes cannot properly reflect user's preference, as the correlation values in this case are dominated by the average ratings of each image. Therefore, we compute the offset (or residual) of a user's score to the ground truth (average) score, and measure correlation between offset values and image attributes instead. We randomly select 8 AMT workers and show the results in Figure 2a, in which dark red color indicates a user prefers these attributes, while dark blur color means the user has relatively lower scores on images with those attributes. It clearly shows that the deviation of each user's ratings (w.r.t. image attributes) are unique. We further visualize example images rated by Worker1 and Worker4 in Figure 2b. As we can see, Worker1 prefers landscape images versus images of crowds while Worker4 prefers images with symmetry attributes over images of people.

To understand the significance of the correlation versus randomness of users' labels, we additionally create two "random " users as the baseline. The two random users




(a)




(b)

Figure 2: (a) Ranking correlation between offsets and image attributes for randomly selected 8 AMT workers and 2 random users. (b) First row: Example images rated by Worker1. The user tends to assign higher scores to landscapes and lower scores to crowd scenes than the average ratings; Second row: Example images rated by Worker4. The user tends to assign higher scores to images containing symmetry attributes but lower scores to people scenes than the average ratings.

are generated by randomly sampling 1,000 images from the training set as their annotated images. The score for each image is set to the ground-truth score (i.e. the average rating of five AMT workers) perturbed by a zero-mean Gaussian random noise with standard deviation of 0.2 and 2, respectively. We choose those standard deviations to simulate two "average" users with a relatively small and a relatively large amount of label offsets deviated from the generic scores. The correlations of their offsets with attributes are also included in Figure 2a, which show no statistically meaningful preference as expected. Compared with "random" users, the correlations on actual users are much stronger, demonstrating that their preferences are indeed related to content and aesthetics attributes instead of random deviations.

For each user, we also compute the sum of the absolute values of the correlation and compare the value with two random users. We run the experiments for 50 times and report the average for the two "random" users. We find all the 111 actual users have higher average correlation scores than the "random" users, showing that the correlations are statistically meaningful.

The analysis demonstrates that score offsets are very effective cues revealing user preferences regarding aesthetics on content and aesthetics attributes. Motivated by this, we derive a novel residual (offset)-based model for learning personalized aesthetics in the next section.

## 5. Approach

In this section, we first introduce our proposed approach to learn personalized aesthetics models by adapting

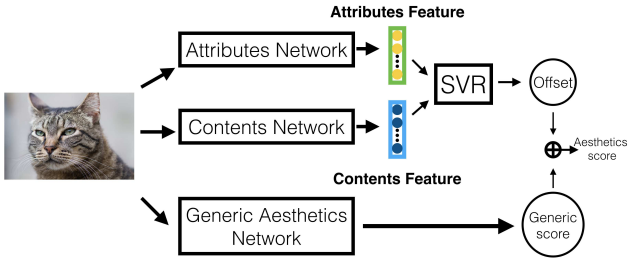



Figure 3: An overview of the proposed residual-based personalized aesthetics model.

a generic aesthetics model to individual users and then continue to derive an active learning algorithm for real-world photo curation applications.

### 5.1. Personalized aesthetics model (PAM)

In real-world photo ranking and curation applications, users often provide a very limited number of aesthetic ratings or feedbacks to their own photos. The lack of labeled exampled makes it difficult to train a meaningful personalized aesthetics model from scratch. Traditional recommendation-based approaches such as collaborate filtering may not be very effective as they require significant overlapping of items rated by different users. In photo curation applications, the user-item matrix could be too sparse to learn effective latent vectors[24].

In order to learn an effective personalized model with good generalization, we aim to capture not only the common aesthetic preference shared across individuals[1] but also the unique aesthetic preference by each individual [29, 33]. Following this idea, we propose to leverage the generic aesthetics model trained to predict the average user's ratings, and focus personalized aesthetics learning on modeling the deviation (residual) of individual aesthetics scores from the generic aesthetics scores. We first train a generic aesthetics model using the FLICKR-AES training set by treating the average rating as the ground truth aesthetic score. Then, given an example set rated by each user, we apply the generic model to each image in the set to compute the residual scores between the user's ratings and the generic scores. Finally, we train a regression model[3] to predict the residual scores. The overview of the approach is illustrated in Figure 3.

**Generic aesthetics prediction** Recent studies[12, 15, 16] have achieved promising results on generic image aesthetics prediction using deep learning. Inspired by these works, we train a deep neural network to predict genetic aesthetic scores. It has the same architecture as in [8] except that we trimmed the number of neurons in the second-to-the-last layer, which we found makes the training more efficient and yields better accuracies. We tested different combinations of loss functions as in [12], but experimentally found that the Euclidean loss function alone works the best on our dataset. We also verified that our generic model achieves results comparable to the recent state of the art by Kong et al. [12] on their AADB dataset [12].

**Residual learning for personalized aesthetics** With the generic aesthetic scores, we can compute residual scores (offsets) for the example images by subtracting them from ratings by each user. Our goal is then reduced to learn a regressor to predict the residual score given any new image.

Due to the lack of annotated examples from each user, training such regressor directly from an image is not practical. Therefore, we propose to use high-level image attributes related to image aesthetics to form a compact feature representation for residual learning. We consider both aesthetic and content attributes inspired by previous studies[5, 12, 15, 17, 20, 22] for automatic assessment of generic aesthetics.

Given the features, we simply use the support vector regressor with a radial basis function kernel to predict the residual score as shown in Equation 1,

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

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 $\nu$ $(0 < \nu \leq 1)$ controls the proportion of the number of support vectors with respect to the number of total training images. We discuss the details of the feature vector $\mathbf{x}_i$ in the following.

**Feature representation for personalized aesthetics** We first leverage the existing AADB[12] dataset which contains around 10,000 images labeled with ten aesthetic attributes[3] to train our aesthetic attribute features. Due to the limited number of training images provided by AADB, we use our trained generic aesthetics network as a pre-trained model, and fine-tune it by multi-task training, i.e. attribute prediction using AADB and aesthetics prediction uisng Flickr-AES. We use the Euclidean loss for both attribute prediction and aesthetics prediction, and fix all the earlier layers of the generic model and only fine-tune its last shared inception

[3]The aesthetics attributes include *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)*

layer and the prediction layers. Given the fine-tuned network, we use the 10-dimensional responses as the aesthetic attributes feature vector $f_{attr}$. Experiments show that the attributes prediction performance of our jointly trained network is significantly better than the one from[12], as shown in Table 1. It demonstrates that features learned from larger-scale aesthetics dataset could also benefit attributes prediction attribute annotations.

As for the content features, we use the off-the-shelf image classification network[8] to extract semantic features (avg pool) from each image. In order to generate compact content attribute features, we first use $k$-means to cluster the images from the FLICKR-AES training set into $k = 10$ semantic categories using the second-to-the-last inception layer output as the feature. We then add a $k$-way softmax layer on top of the network and fine-tune the layer with the cross-entropy loss. The 10-dimensional outputs of the network are defined as the content attributes feature vector $f_{cont}$. We concatenate the two feature vectors $\mathbf{x} = [f_{attr}, f_{cont}]^T$, to form our final feature representation to personalized aesthetics learning. Experiments show the concatenation of attributes and content features achieve better results than using each of them alone.

**Algorithm 1:** Active-PAM

**Input:** Unrated photo set $\mathcal{N} = \{p_i\}_{i=1,K}$, the aesthetic feature vectors $v_i$, the maximum number of example ratings $m$;

Initialize the set of rated examples as: $\mathcal{R} = \emptyset$ ;

Randomly select a subset $\mathcal{S}$ of $\lfloor K/10 \rfloor$ images from $\mathcal{N}$ and move them to $\mathcal{R}$: $\mathcal{N} = \mathcal{N} \setminus \mathcal{S}, \mathcal{R} = \mathcal{R} \cup S$ ;

**while** $|\mathcal{R}| < m$ **do**

  Train a regressor to predict the residual score $\{r_i\}$;

  Calculate the weight for each annotated image

$$w_i = (1 - \frac{|r_i|}{\sum_{i=1}^{|\mathcal{R}|}(|r_i|)}), p_i \in \mathcal{R} \quad (2)$$

  Find $p_q$ that

$$\max_q \sum_{j=1}^{|\mathcal{R}|} w_j \text{dist}(v_q, v_j), p_q \in \mathcal{N}, p_j \in \mathcal{R} \quad (3)$$

  Add the selected image to $R$ and update $N$: $\mathcal{R} = \mathcal{R} \cup \{p_q\}$ and $\mathcal{N} \setminus p_q$ ;

### 5.2. Active personalized image aesthetics learning (Active-PAM)

In real-world applications such as interactive photo curation, users can continuously provide ratings regarding their aesthetics preference during the photo selection and ranking process[33, 34]. Instead of waiting for users to provide ratings on arbitrary images, we can use active learning to automatically select the most representative images for users to rate, and learn from their feedback online. To minimize the user effort, we propose a new active learning algorithm to optimize sequential selection of training images for personalized aesthetics learning. Specifically, we consider the following two criteria: 1) the selected images should cover diverse aesthetic styles as quickly as possible with minimum redundancy; 2) the images with large residual scores between user's ratings and the generic aesthetics scores are more informative.

Based on these criteria, we design our active selection algorithm as follows. For each image $p_i$ in the collection $\mathcal{N}$, we denote its aesthetics score predicted by the generic aesthetics network as $s_i$, features extracted at the second-to-the-last layer output as $f_i$. The aesthetic feature capturing the aesthetic styles of the image can then be represented as $v_i = [w_a f_i, s_i]$, where $w_a$ is a constant balancing the two terms. We can then measure the distance between any two images $p_i$ and $p_j$ using the Euclidean distance $\text{dist}(v_i, v_j)$.

Given a set of images $\mathcal{R}$ already 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 $\mathcal{R}$, $d_i = \sum_{j=1}^{|\mathcal{R}|} \text{dist}(v_i, v_j)$, $p_j \in \mathcal{R}$. At each step, we can choose the image with the largest $d_i$ according to the first criterion. In order to incorporate the second criterion at the same time, we can encourage selecting the image producing large residuals in $\mathcal{R}$. We denote the residual score as $r_j$ and assign weight $w_j$ to each image using Equation 2. We apply the weights to the overall distance, resulting in Equation 3. The details of the active learning algorithm are described in Algorithm 1.

| | 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** |

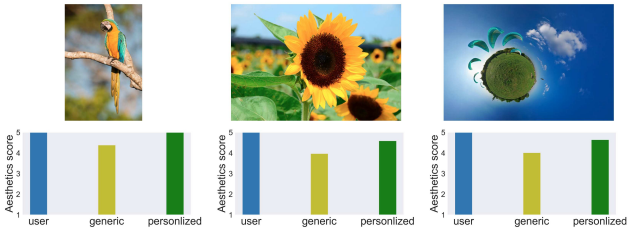
Table 1: Attributes comparison of the baseline[12] and our results. Both calculated by correlation $\rho$. Jointly training attributes and aesthetics improves the attributes prediction.

## 6. Experiments

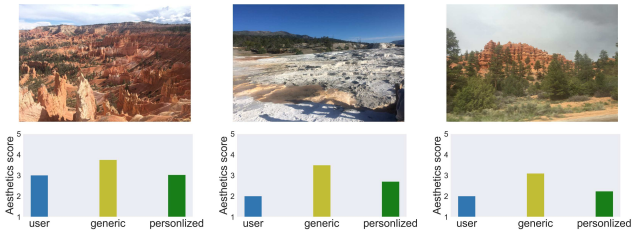
In this section, we present the experimental evaluation of our personalized aesthetics model (PAM) and our active learning approach (Active-PAM) on both the FLICKR-AES and the REAL-CUR datasets. Figure 4 introduces visual examples to show how the model works.

### 6.1. Implementation details

The earlier layers of our generic aesthetics network are initialized from the Inception-BN network [8], whereas the last trimmed inception module is randomly initialized using




(a) One example user from FLICKR-AES




(b) One example user from REAL-CUR

Figure 4: Example results of personalized aesthetics prediction from two users. The examples in (a) comes from FLICKR-AES and the examples in (b) comes from REAL-CUR. The blue bar is the user rating, the yellow bar is the generic aesthetics prediction and the green bar is the personalized aesthetics prediction for this user. As can be seen, our personalized model more accurately predicts user ratings than the generic model.

"Xavier"[6]. The network is then fine-tuned on FLICKR-AES. During training, images are warped to $256 \times 256$ and then randomly cropped to $224 \times 224$ before feeding into the network. The learning rate is initialized as 0.001 and periodically annealed by 0.96. The weight decay is 0.0002 and the momentum is 0.9.

### 6.2. Evaluation on personalized aesthetics models

To evaluate our personalized aesthetics model (PAM), we compare PAM with a baseline model as well as a commonly used collaborate filtering approach and show that PAM works significantly better than the other two. We further analyze the effectiveness of content and aesthetic attribute features used in our model.

**Comparison with other methods.** We compare PAM with two other methods: 1) a baseline SVM regressor that directly predicts user-specific aesthetics scores on test images based on user-provided training images, and 2) a commonly used collaborate filtering approach, non-linear Feature-based Matrix Factorization (FPMF), which achieves better performance than other matrix factorization approaches on individual color aesthetics recommendation[24]. In both methods, we use the same content and aesthetic attribute features as in our method.

We evaluate these three methods on the test workers in FLICKR-AES. During evaluation, for each test worker, we randomly sample $k$ images from the ones he or she labeled, and use them as training images. All the remaining images are then used for testing. For the non-linear FPMF, all the other workers in the FLICKR-AES training set are also included for training. Due to the randomness of training image selection, we run the experiments 50 times for each test worker, and report the averaged results as well as the standard deviation.

Following [12], ranking correlations are used to measure the consistency between the predictions and the ground-truth user scores. The mean ranking correlation of the generic aesthetics model over all the test workers are 0.514. In Table 2, we show the improvement in terms of correlation for each method compared with the generic model, with $k = 10$ and $k = 100$, respectively. We can see the SVM model that directly predict scores does not work on this problem, as its results are even worse than the ones from the generic model. It tries to directly learn each user's flavor regarding aesthetics from very limited data without considering generic aesthetics understanding, which is accordingly very unstable and hardly generalizable. By contrast, our method (last row in Table 2) works even with 10 training images, and has much more significant improvement with more training examples. It validates the design of our residual-based model, which fully leverages the common understanding of aesthetics existing in the generic aesthetics network, and focuses on the score offsets that directly correspond to users' unique preference compared with generic aesthetics. Our model also significantly outperforms FPMF, which only has marginal improvement even when using 100 training images.

**Ablation study on features.** We also show the results of PAM when trained only using the content feature or only using the aesthetics attribute feature, respectively, in the 3rd and 4th row in Table 2. We can see that both content and aesthetics attributes can be used to model personalized aesthetics, as the correlations with users' ground truth scores also increase when using more user provided training images. Nevertheless, using both features gives the best performance. It further demonstrates that users' preference on image aesthetics are related to both image content and aesthetic attributes.

### 6.3. Evaluation on active learning

**Comparison with other methods.** We compare our method with three other active learning methods: 1) Greedy[35], which selects the sample that has the minimum largest distance to the already selected samples at every iteration; 2) MCAL[4], which chooses samples by clustering the candidate images to be selected and selected images that are not support vectors; 3) Query by Committee (QBC) [2],

which generates a committee of models by using ensemble algorithms. In our experiment, we generate 5 committees using Bagging for QBC [19]. We choose these three methods for comparison because they deal with regression problem, whereas classic active learning approaches[4, 28] for classification are not applicable here, as their criteria such as margin-based sampling are not suitable for continuous aesthetics scores. In addition, we also add another baseline where all the images are randomly selected.

To evaluate an active learning method, we start with 10 randomly selected images to train the initial PAM. Based on the initial results, we keep selecting new images from user's photo album using the active learning methods and updating the PAM model, until the number of selected images reaches 100. Different methods may select different images for model update. To compare these methods on the same test images, we chose to use the entire photo album for evaluation. It is also consistent with the real application scenario of personalized aesthetics, where the algorithm is able to access and actively select any image in the photo album, and the quality of overall ranking on all images in the album is the most important factor for the user.

Due to the randomness in the initialization of the PAM model, we repeat the experiments 10 times for each method. The average performance is shown in Figure 5 and the standard deviation for all methods is less than 0.001. Our active selection method outperforms all the other baseline methods as well as random selection.

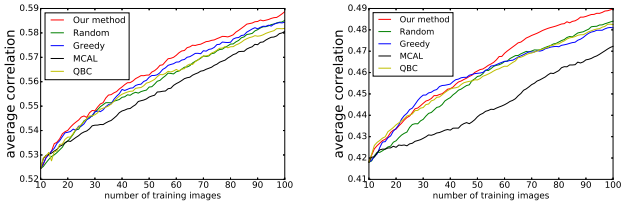
To further examine the generalizablility of the PAM models updated by different active learning approaches, instead of evaluating on the entire photo album, we remove all the images that have already been selected for model update, and only evaluate on the remaining images the model has not seen before. We note that it is not a totally fair comparison, as images used for evaluation may be different for different active learning approaches, due to the different images they selected for model update. But it still gives us a sense how the model works on new images. The results are reported in Figure 6. When evaluating on those unseen images, the PAM model updated by our active selection performs significantly better.

## 7. Conclusion

In this work, we address the problem of personalized image aesthetics, and introduce two new datasets to facilitate investigatio of this problem. We propose a novel residual-based personalized aesthetics model for accommodating individual aesthetics taste with limited annotated examples. 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 research direction is to investigate additional cues such as content redundancy, image quality or face recognition for improving user experience in real-world applications.

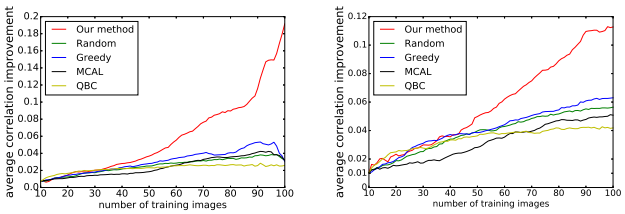
| | 10 images | 100 images |
|---|---|---|
| Direct score prediction | -0.352 $\pm$ 0.050 | -0.176 $\pm$ 0.064 |
| FPMF (only attribute) | -0.003 $\pm$ 0.004 | 0.002 $\pm$ 0.003 |
| FPMF (only content) | -0.002 $\pm$ 0.002 | 0.002 $\pm$ 0.010 |
| FPMF (content and attribute) | -0.001 $\pm$ 0.003 | 0.010 $\pm$ 0.007 |
| PAM (only attribute) | 0.004 $\pm$ 0.003 | 0.025 $\pm$ 0.013 |
| PAM (only content) | 0.001 $\pm$ 0.004 | 0.021 $\pm$ 0.017 |
| PAM (content and attribute) | **0.006** $\pm$ **0.003** | **0.039** $\pm$ **0.012** |

Table 2: Comparison with direct score prediction (using SVM), non-linear FPMF[24], PAM with only content feature, PAM with only attribute feature and PAM with both features using different number of training images from each worker. The results are average correlation improvement.




(a) Results on FLICKR-AES (b) Results on REAL-CUR.

Figure 5: Evaluation of active learning approaches on (a) FLICKR-AES and (b) REAL-CUR on all the images in the photo album.




(a) Results on FLICKR-AES (b) Results on REAL-CUR

Figure 6: Evaluation of active learning approaches on (a) FLICKR-AES and (b) REAL-CUR on unseen images in the photo album.

## 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, 5

[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. 3, 7

[3] C.-C. Chang and C.-J. Lin. Training nu-support vector regression: theory and algorithms. *Neural Computation*, 14(8):1959–1978, 2002. 5

[4] B. Demir and L. Bruzzone. A multiple criteria active learning method for support vector regression. *Pattern recognition*, 47(7):2558–2567, 2014. 3, 7

[5] 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, 3, 5

[6] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010. 7

[7] R. He and J. McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web*, pages 507–517. International World Wide Web Conferences Steering Committee, 2016. 2

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

[9] 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. 2

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

[11] 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

[12] 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, 5, 6, 7

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

[14] 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

[15] 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. 1, 2, 5

[16] 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, 5

[17] 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, 3, 5

[18] 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

[19] 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.

[20] 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, 5

[21] 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

[22] 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. 2, 5

[23] J. L. Myers, A. Well, and R. F. Lorch. *Research design and statistical analysis*. Routledge, 2010. 3

[24] 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, 5, 7, 8

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

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

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

[28] 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

[29] 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, 5

[30] S. Wang, Y. Wang, J. Tang, K. Shu, S. Ranganath, and H. Liu. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In *Proceedings of the 26th International Conference on World Wide Web*, pages 391–400. International World Wide Web Conferences Steering Committee, 2017. 1

[31] Z. Wang, F. Dolcos, D. Beck, S. Chang, and T. S. Huang. Brain-inspired deep networks for image aesthetics assessment. *arXiv preprint arXiv:1601.04155*, 2016. 2

[32] R. Willett, R. Nowak, and R. M. Castro. Faster rates in regression via active learning. In *Advances in Neural Information Processing Systems*, pages 179–186, 2005. 3

[33] C.-H. Yeh, B. A. Barsky, and M. Ouhyoung. Personalized photograph ranking and selection system considering positive and negative user feedback. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 10(4):36, 2014. 1, 2, 5, 6

[34] C.-H. Yeh, Y.-C. Ho, B. A. Barsky, and M. Ouhyoung. Personalized photograph ranking and selection system. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 211–220. ACM, 2010. 1, 2, 6

[35] H. Yu and S. Kim. Passive sampling for regression. In *2010 IEEE International Conference on Data Mining*, pages 1151–1156. IEEE, 2010. 7