

Article

# Personalized Image Aesthetics Assessment via Multi-Attribute Interactive Reasoning

Hancheng Zhu <sup>1</sup>, Yong Zhou <sup>1,\*</sup>, Zhiwen Shao <sup>1</sup>, Wenliang Du <sup>1</sup>, Guangcheng Wang <sup>2</sup> and Qiaoyue Li <sup>3</sup>

<sup>1</sup> School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

<sup>2</sup> School of Transportation and Civil Engineering, Nantong University, Nantong 226019, China

<sup>3</sup> Department of Optoelectronics and Energy Engineering, Suzhou City University, Suzhou 215104, China

\* Correspondence: yzhou@cumt.edu.cn

**Abstract:** Due to the subjective nature of people's aesthetic experiences with respect to images, personalized image aesthetics assessment (PIAA), which can simulate the aesthetic experiences of individual users to estimate images, has received extensive attention from researchers in the computational intelligence and computer vision communities. Existing PIAA models are usually built on prior knowledge that directly learns the generic aesthetic results of images from most people or the personalized aesthetic results of images from a large number of individuals. However, the learned prior knowledge ignores the mutual influence of the multiple attributes of images and users in their personalized aesthetic experiences. To this end, this paper proposes a personalized image aesthetics assessment method via multi-attribute interactive reasoning. Different from existing PIAA models, the multi-attribute interaction constructed from both images and users is used as more effective prior knowledge. First, we designed a generic aesthetics extraction module from the perspective of images to obtain the aesthetic score distribution and multiple objective attributes of images rated by most users. Then, we propose a multi-attribute interactive reasoning network from the perspective of users. By interacting multiple subjective attributes of users with multiple objective attributes of images, we fused the obtained multi-attribute interactive features and aesthetic score distribution to predict personalized aesthetic scores. Experimental results on multiple PIAA datasets demonstrated our method outperformed state-of-the-art PIAA methods.

**Keywords:** image aesthetics assessment; personalized aesthetic experiences; multiple attributes; interactive reasoning

**MSC:** 68U10; 68T05



**Citation:** Zhu, H.; Zhou, Y.; Shao, Z.; Du, W.; Wang, G.; Li, Q. Personalized Image Aesthetics Assessment via Multi-Attribute Interactive Reasoning. *Mathematics* **2022**, *10*, 4181. <https://doi.org/10.3390/math10224181>

Academic Editors: Grigoreta-Sofia Cojocar, Adriana-Mihaela Guran and Laura-Silvia Dioşan

Received: 8 October 2022

Accepted: 7 November 2022

Published: 9 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

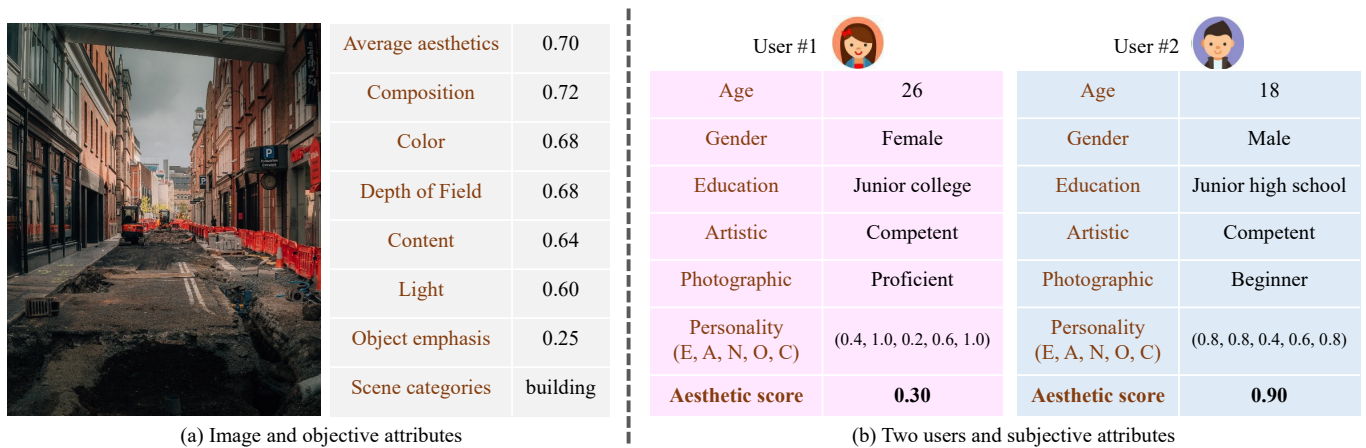
In the past few years, with the prevalence of social networks (such as Facebook and Wechat), people usually use multimedia data such as images to obtain information and for other visual needs. Therefore, the visual experience of providing images in these social networks plays a key role in attracting users. In this context, it is desirable to develop image aesthetics assessment (IAA), which can simulate users' visual experiences and automatically assesses the aesthetics of images, e.g., digital cameras provide users with aesthetic evaluation suggestions when taking photos. Consequently, massive IAA methods [1,2] have been proposed by researchers in the pattern recognition and computer vision communities, which has applicable value for various tasks, e.g., photo retrieval [3], image management [4], image enhancement [5], image synthesis [6], and image recommendation [7].

Typically, IAA approaches can be classified into two broad categories: generic image aesthetics assessment (GIAA) and personalized image aesthetics assessment (PIAA) [8]. As the name indicates, GIAA aims to infer the aesthetic experiences perceived by most

people [9], whereas PIAA is designed for the aesthetic ratings of a certain individual user for images [10]. Early IAA methods mainly leveraged general attributes in photography and artistic painting (e.g., composition, color, light) to measure the aesthetics of images for most people (GIAA) [2]. Specifically, the average rating of an image annotated by different people is used as the “ground truth” to classify the image into high and low aesthetic categories [11,12] or to map the image to a certain aesthetic score [13,14]. However, these average results ignore an important fact that people’s aesthetic experience of images is subjective. In view of this, existing GIAA methods mainly focus on directly predicting the aesthetic distribution of most people’s ratings on images [15–18]. Although the image aesthetic distribution can reflect the aesthetic subjectivity of people to a certain extent, this task only measures people’s aesthetic ratings from the perspective of images. All in all, GIAA methods are unable to infer individual users’ aesthetic preferences for images, which is very valuable in many user-centric applications (e.g., personalized image recommendation [19], personalized image captioning [20] and personalized image enhancement [21]). To deal with this issue, PIAA is proposed to obtain individual users’ personalized aesthetic experience of images [22].

PIAA is a user-oriented approach that can only utilize the annotated data provided by each individual user to build a PIAA model [10]. Usually, the number of annotated data provided by a user is limited, which is unable to directly train an efficient PIAA model based on a deep learning framework. Consequently, existing PIAA models mainly rely on a large amount of labeled data rated by most users to train a prior model and further use the labeled data of individual users for model fine-tuning [22–29]. These prior models can be summarized into two types, including learning from the generic aesthetic results of most users [22–26] or learning from the personalized aesthetic results of a large number of individual users [27–29]. The former prior model can capture generic aesthetic experience from the perspective of images, while the latter prior model directly obtains personalized aesthetic experience from the perspective of users. However, the prior knowledge obtained from images eliminates the aesthetic differences among individual users, while the prior knowledge achieved directly from individual users cannot efficiently capture the general aesthetics of images.

To alleviate the above issues, the prior model for PIAA should not only learn the general aesthetics of images, but also model the aesthetic differences of individual users. Specifically, the general aesthetics of an image is usually determined by its objective attributes [30]. For example, Figure 1a shows an image and the corresponding objective attributes. We can observe that the generic (average) aesthetics is closely related to multiple attributes, and these attributes jointly influence most users’ aesthetic experience of images. Besides, the aesthetic differences among individual users are usually affected by their own subjective attributes [10]. As shown in Figure 1b, the subjective attributes of the two users are quite different. For instance, User #1 has better education and photography skills than User #2, which makes User #1 more stringent about the attributes such as composition and light of the image. User #2 may prefer scenes such as buildings. All in all, User #1 gives the image a lower aesthetic score (0.3), while User #2 has a higher aesthetic score for the image (0.9). Therefore, exploring the close relationship between the multiple objective attributes of images and the multiple subjective attributes of users is the premise of inferring the personalized aesthetics of a specific user. However, this interactive relationship has not been exploited in the prior model of existing PIAA methods [22–29]. To this end, we can leverage the interactive relationship between subjective and objective attributes to capture aesthetic prior knowledge (multi-attribute interactions). Even when a user provides limited annotated data, the aesthetic prior knowledge can also stably utilize the relationship between the subjective attributes of similar users and the objective attributes of images for reasoning about the user’s aesthetic preferences.



**Figure 1.** An image and two different users who rated it from the Personalized image Aesthetics database with Rich Attributes (PAPA) [10]. To their right, some objective attributes of the image and several subjective attributes of the users are shown. These numerical attributes and aesthetic scores are normalized between 0 and 1, and higher values indicate stronger attributes and aesthetics. The personality of users here is measured by the Big-Five traits (extroversion (E), agreeableness (A), neuroticism (N), openness (O), and conscientiousness (C)) [31].

In this paper, we propose a personalized image aesthetics assessment method via multi-attribute interactive reasoning (PIAA-MIR). In order to reveal the personalized aesthetic preference of users for images, we expect to capture the aesthetic prior model that reflects the potential interaction between the subjective attributes of users and the objective attributes of images. Compared with the existing prior models that only learn the generic aesthetics of most users [22–26] or the personalized aesthetics of a large number of individual users [27–29], the proposed multi-attribute interaction can effectively characterize the aesthetic mutual influence of users and images to accurately evaluate personalized aesthetic preferences. Specifically, we first propose a generic aesthetics extraction module from the perspective of images to simultaneously predict multiple objective attributes and aesthetic distributions of images. From the perspective of users, a multi-attribute interaction reasoning network is then introduced to capture the interaction between multiple attributes of users and images. To obtain the multi-attribute interaction, we utilized the outer-product [32] to calculate the pairwise correlations between multiple attributes of users and images. Based on the multi-attribute interactive features and aesthetic score distribution, we used a regressor to fuse them for obtaining personalized aesthetic scores. To sum up, the main contributions of the proposed method are as follows:

- We excavated the fundamental factors of users’ personalized aesthetic preferences for images by constructing a multi-attribute interaction, which alleviates the insufficient problem of directly obtaining prior knowledge only from the generic aesthetics of images or the personalized aesthetics of a large number of individual users.
- We propose a generic aesthetics extraction module that can simultaneously predict multiple attributes and aesthetic distributions of images. In the multi-attribute interactive reasoning network, we can not only leverage multiple attributes of images and users to construct an effective interaction, but also further use the multi-attribute interactive features and aesthetic score distribution to jointly model personalized aesthetic scores.
- We propose a personalized image aesthetics assessment method via multi-attribute interactive reasoning (PIAA-MIR), whose experimental results on several PIAA databases demonstrated that the proposed PIAA-MIR outperformed state-of-the-art PIAA methods. Besides, ablation studies also showed the effectiveness of our method in learning a personalized aesthetic prior model.

## 2. Related Works

Since existing PIAA methods are mainly built on the GIAA model, we first review some works related to the GIAA methods and then introduce the PIAA methods.

### 2.1. Generic Image Aesthetics Assessment

Early researchers believed that people had a consensus on the aesthetic experience of images [33] and generic image aesthetics perceived by most people could be measured by aesthetic rules in photography (e.g., light, colorfulness, and composition) [34]. Generally, GIAA methods can be divided into three categories: aesthetic binary classification [11,12], aesthetic score regression [13,14], and aesthetic distribution prediction [15–18]. The goal of the aesthetic binary classification task is to classify images into “high” and “low” categories according to the aesthetic ratings of most people. Specifically, Murray et al. [11] introduced a large general-purpose IAA database, AVA, and utilized hand-crafted features to train an SVM for image aesthetic classification. Compared with the aesthetic binary classification task, aesthetic score regression needs to more accurately predict image aesthetic scores. For instance, Kong et al. [13] employed a deep Siamese network based on image pair ranking learning, which can simultaneously predict the aesthetic attributes and content of images and further learn to rank the aesthetic scores of images on the basis of aesthetic attributes and content information.

Regardless of aesthetic binary classification or aesthetic score regression, it is necessary to process the aesthetic ratings of different people into a unified result (“high” or “low” and aesthetic score), which will introduce label uncertainty to a certain extent. The main reason is that people’s aesthetic experiences are highly subjective, which makes the unified result unable to effectively describe the image aesthetics perceived by different people. Therefore, the task of aesthetic distribution prediction that directly models the image score distribution rated by most people has received great attention from researchers. For example, Talebi et al. [2] used the earth mover’s distance (EMD) loss function to train an IAA model for predicting the image aesthetic distribution. The above methods mainly focus on the image aesthetic distribution, ignoring the intrinsic relationship among the three tasks of image aesthetics binary classification, aesthetic score regression, and aesthetic distribution prediction. Therefore, some recent studies have proposed a unified deep learning framework for the three GIAA tasks [15–17]. For example, Zeng et al. [16] proposed a deep model with a unified probabilistic formula and introduced a loss function that is effective for all three GIAA tasks to optimize the deep model. Based on the above analysis, we can find that the current GIAA research mainly focuses on aesthetic distribution prediction. Therefore, our generic aesthetics extraction module exploits the score distribution to represent generic image aesthetics.

### 2.2. Personalized Image Aesthetics Assessment

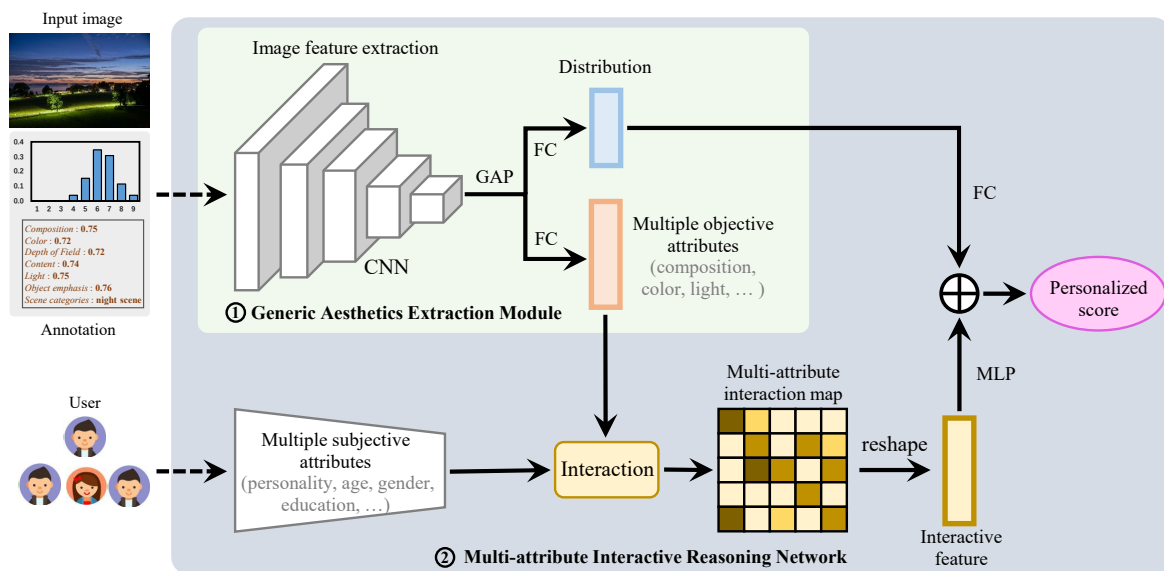
The purpose of PIAA is to evaluate images by simulating the visual aesthetics of individual users [22]. Since users’ aesthetic preferences are affected by multiple factors such as age, education, and behavioral habits [35,36], the PIAA for a specific user is more complicated and difficult than the GIAA for generic users. Due to the limited labeled samples provided by individuals, PIAA is a small sample learning task [28]. Existing PIAA models are usually built on a prior model with generic aesthetic knowledge, which utilizes aesthetic data annotated by massive users for model training [22–29].

Among them, one approach is to take generic aesthetic results rated by most users as the target for prior model learning from the perspective of images. For instance, Ren et al. [22] found that users’ aesthetic preferences were closely related to image content and aesthetic attributes and leveraged the average aesthetic scores of images, image content, and aesthetic attributes to jointly infer personalized aesthetic scores. Li et al. [8] built a prior model for the PIAA task by using the aesthetic distribution of images and the Big-Five personality traits of users who prefer these images. However, these prior models learned from images eliminate the aesthetic differences among individual users.

Another approach is to learn the prior model directly from the personalized aesthetic results of a large number of individual users from the perspective of users. In [28], Zhu et al. proposed a PIAA model based on bi-level gradient optimization meta-learning, which directly captured an aesthetic prior knowledge by training the PIAA tasks of extensive users. Hou et al. [29] trained a prior aesthetic pattern for all individual users by leveraging the interaction between user preferences and image content. In [27], the authors inferred the Big-Five traits of users from their rated images and used the personalized aesthetics of massive individual users to train a prior model. However, the above aesthetic prior models learned directly from individual users are inefficient in capturing the general aesthetics of images. To this end, we expect that the prior model of PIAA can both learn the general aesthetics of images and model the aesthetic differences of various users. Therefore, we utilized multiple attributes of images and users to characterize general aesthetics and aesthetic differences, respectively, and capture the stable interactive relationship between objective attributes and subjective attributes for easily inferring users' personalized aesthetics of images.

### 3. Proposed Method

This section introduces the personalized image aesthetics assessment method via multi-attribute interactive reasoning, which is called PIAA-MIR. In the proposed PIAA-MIR, we obtain the prior model for the PIAA task of an individual user by implementing a multi-attribute interaction between users and images. Figure 2 shows the overview architecture of our PIAA-MIR, whose training process can be divided into three steps. In the first step, a generic aesthetics extraction module is the software command line to proceed with the extraction, which is trained with images and the annotated multiple aesthetic attributes and score distribution. In the second step, we built a prior model from a multi-attribute interaction between users, as well as their rated images, which further reasons personalized scores by fusing interactive features and the score distribution. In the third step, we leveraged an individual user's personalized aesthetic data to fine-tune the prior model for obtaining the PIAA model of the user.



**Figure 2.** The overview architecture of our PIAA model, whose training process can be divided into three steps. In the first step, a generic aesthetics extraction module is used to simultaneously predict multiple attributes and aesthetic distributions from the perspective of images. In the second step, a multi-attribute interaction reasoning network is then introduced from the perspective of users to capture a multi-attribute interaction between users, as well as their rated images. Based on interactive features and the score distribution, a prior model is built to fuse them for obtaining personalized aesthetic scores. In the third step, the PIAA model of an individual user can be obtained by fine-tuning the user's personalized data.



### 3.1. Generic Aesthetics Extraction

To obtain the generic aesthetics of images, we designed a generic aesthetic extraction module to jointly infer multiple objective attributes and the aesthetic score distribution. Consequently, we introduced a convolutional-neural-network (CNN)-based multi-task learning [8,37] to extract the shared image features of generic aesthetic attributes and the distribution. The proposed CNN was inherited from the typical ResNet [38], which removes the full connection layer. As shown in the upper part of Figure 2, we adopted the CNN parameters pre-trained on ImageNet [39] as initial weights, which further use a global average pooling (GAP) and two fully connected layers (FC) for mapping images to multiple aesthetic attributes and score distributions.

In particular, for an input image  $x$ , the generic aesthetic extraction module can be formulated as

$$\hat{\mathbf{a}} = FC_{\theta_a}(GAP(f_{\theta}(x))), \hat{\mathbf{d}} = FC_{\theta_d}(GAP(f_{\theta}(x))), \tag{1}$$

where  $\theta$  represents the CNN parameters  $f_{\theta}$  and  $FC_{\theta_a}$  and  $FC_{\theta_d}$  denote the parameters of an FC layer corresponding to the predicted multiple aesthetic attributes  $\hat{\mathbf{a}}$  and score distribution  $\hat{\mathbf{d}}$ , respectively. Since this module aims to extract the aesthetic attributes and score distributions of images rated by most people, we assumed  $\mathcal{D}_{img} = \{x_i, \mathbf{a}_i, \mathbf{d}_i\}_{i=1}^{N_a}$  as the set for training the generic aesthetic extraction module, where  $\mathbf{a}_i$  and  $\mathbf{d}_i$  indicate some annotated aesthetic attributes and score distribution of the  $i$ -th image  $x_i$  ( $i = 1, 2, 3, \dots, N_a$ ), respectively. Besides,  $N_a$  denotes the number of images in this training set.

To enable the proposed generic aesthetic extraction module to effectively predict aesthetic attributes and the score distribution, we employed the earth mover’s distance (EMD) [40] and  $l_2$  loss functions to jointly optimize the parameters of this module ( $\theta, \theta_a$ , and  $\theta_d$ ), which is defined as

$$\mathcal{L}_a = \frac{1}{N_a} \sum_{i=1}^{N_a} \left( (\mathbf{a}_i - \hat{\mathbf{a}}_i)^2 + \left( \frac{1}{P} \sum_{k=1}^P |CDF_{\mathbf{d}_i}(k) - CDF_{\hat{\mathbf{d}}_i}(k)|^2 \right)^{\frac{1}{2}} \right), \tag{2}$$

where  $\hat{\mathbf{a}}_i$  and  $\hat{\mathbf{d}}_i$  are the predicted aesthetic attributes and score distribution by feeding the  $i$ -th image into the generic aesthetic extraction module. Similar to [2], classes in the image aesthetic score distribution are inherently ordered as  $d_i^{p_1} < \dots < d_i^{p_p}$ . Therefore, the EMD, which contains the cumulative distribution function (CDF), is sensitive to the order of aesthetic score buckets, which is suitable for calculating the loss of the image aesthetic distribution. Specifically,  $P$  indicates the number of aesthetic score buckets, and  $CDF_{\mathbf{d}_i}(k) = \sum_{j=1}^k d_i^{p_j}$  represents the cumulative distribution function, where  $d_i^{p_j}$  denotes the probability of the  $j$ -th score bucket and  $\sum_{j=1}^P d_i^{p_j} = 1$ . In this way, the generic aesthetics extraction module that can simultaneously predict multiple attributes and aesthetic distributions of images can be obtained by using the training data of  $\mathcal{D}_{img}$  from the perspective of images.

### 3.2. Multi-Attribute Interaction Reasoning

Before building the multi-attribute interaction, we need to utilize multiple subjective attributes to characterize individual users. Assume that  $\mathbf{s}$  represents the subjective attributes of an individual user, which can be collected by users answering several questionnaires [10]. To enable the prior model also to robustly capture personalized aesthetic differences from the perspective of users, we leveraged a large number of users’ personalized aesthetic data on images to train the multi-attribute interactive inference network.

Suppose that  $\mathcal{D}_{users} = \{\mathbf{s}_j, \{x_{i,j}, y_{i,j}\}_{i=1}^{N_s}\}_{j=1}^{N_b}$  denotes the set for training the multi-attribute interactive inference network, where  $\mathbf{s}_j$  represents some subjective attributes of the  $j$ -th user ( $j = 1, 2, 3, \dots, N_b$ ) and  $y_{i,j}$  indicates the user’s personalized score for the image  $x_{i,j}$  ( $i = 1, 2, 3, \dots, N_s$ ). For the image  $x_{i,j}$ , the multiple objective attributes and score

distribution can be extracted from the trained generic aesthetics extraction module, which is formulated as

$$\hat{\mathbf{a}}_{i,j} = FC_{\theta_a}(GAP(f_{\theta}(x_{i,j}))), \hat{\mathbf{d}}_{i,j} = FC_{\theta_d}(GAP(f_{\theta}(x_{i,j}))), \tag{3}$$

where  $\hat{\mathbf{a}}_{i,j}$  and  $\hat{\mathbf{d}}_{i,j}$  are the predicted aesthetic attributes and score distribution of the  $i$ -th image in the subset of the  $j$ -th user. As shown in Figure 1, since the user’s personalized aesthetic preference for images is affected by multiple attributes from both sides, we need to obtain all pairwise interactive relationships between subjective attributes and objective attributes. To achieve this, we employed the outer-product [32] to obtain the pairwise interactions between multiple attributes of users and images, which takes the form

$$\mathcal{A}_{i,j} = \mathbf{s}_j \otimes \hat{\mathbf{a}}_{i,j}, \tag{4}$$

where  $\mathcal{A}_{i,j} \in \mathbb{R}^{d_s \times d_a}$  denotes the multi-attribute interaction map,  $\mathbf{s}_j \in \mathbb{R}^{d_s \times 1}$  represents the attributes of the  $j$ -th user,  $\hat{\mathbf{a}}_{i,j} \in \mathbb{R}^{d_a \times 1}$  represents the attributes of the  $i$ -th rated image, and  $\otimes$  is the operation of the outer-product. In addition,  $d_s$  and  $d_a$  indicate the number of users’ subjective attributes and image objective attributes, respectively. The elements in the interaction map  $\mathcal{A}_{i,j}$  reflect the aesthetic preferences of users’ subjective attributes to image objective attributes at different dimensions. For example, if a testing user has similar subjective attributes to some trained users, his/her aesthetic preference for images can be inferred from the stable relationships learned from the multi-attribute interaction map.

To make the prior model learn the aesthetic differences among individual users, we further used the interaction map for reasoning users’ personalized aesthetic scores for images. For this purpose, the interaction map  $\mathcal{A}_{i,j}$  was reshaped to an interactive feature  $\mathbf{I}_{i,j}$ , and we leveraged a two-layer multilayer perceptron (MLP) to map the interactive feature into aesthetic difference scores between different users, which is given by

$$\hat{r}_{i,j} = MLP_{\theta_r}(\mathbf{I}_{i,j}), \tag{5}$$

where  $\hat{r}_{i,j}$  denotes the aesthetic difference score of the  $j$ -th user for the  $i$ -th image relative to most users and  $\theta_r$  indicates the parameters of  $MLP_{\theta_r}$ , which contains two FC layers. As mentioned above, the generic aesthetics of images can also affect users’ personalized aesthetic preferences. Instead of taking the average ratings as the generic scores [8,27], we utilized an FC layer to fuse the score distribution and aesthetic difference score for obtaining a personalized score, which can be formulated as

$$\hat{y}_{i,j} = FC_{\theta_s}(\hat{\mathbf{d}}_{i,j}) + \hat{r}_{i,j}, \tag{6}$$

where  $\hat{y}_{i,j}$  indicates the predicted aesthetic score and  $\theta_s$  denotes the parameters of  $FC_{\theta_s}$ . Then, we employed the  $l_2$  loss function to optimize the parameters of the MLP and FC layers ( $\theta_r$  and  $\theta_s$ ), which is defined as

$$\mathcal{L}_s = \frac{1}{N_b} \sum_{j=1}^{N_b} \left( \frac{1}{N_s} \sum_{i=1}^{N_s} (y_{i,j} - \hat{y}_{i,j})^2 \right), \tag{7}$$

where  $N_b$  and  $N_s$  represent the number of training users and the corresponding rated images, respectively. In this way, the proposed prior model can capture a robust multi-attribute interaction map by learning extensive users’ personalized aesthetic ratings of images from the perspective of users. Based on the learned multi-attribute interaction, the proposed prior model can be efficiently transferred to the personalized aesthetics of a target user through fine-tuning a small number of user-specific aesthetic data.

### 3.3. PIAA Fine-Tuning for a Specific User

Since PIAA is aimed at the aesthetic preferences of a specific individual user, we leveraged a user's personalized aesthetic data to fine-tune the prior model for obtaining the PIAA model. Assume that  $\mathcal{D}_u = \{s_u, \{x_i^u, y_i^u\}_{i=1}^{N_s}\}$  represents the training set of a specific user, where  $N_s$  denotes the number of small samples annotated by the user and  $s_u$  denotes subjective attributes. Besides,  $x_i^u$  and  $y_i^u$  represent the  $i$ -th image, as well as the corresponding aesthetic score. Firstly, we leveraged the generic aesthetics extraction module to obtain the objective attributes and score distribution of the image, which can be defined as

$$\hat{a}_i^u = FC_{\theta_a}(GAP(f_{\theta}(x_i^u))), \hat{d}_i^u = FC_{\theta_d}(GAP(f_{\theta}(x_i^u))), \quad (8)$$

where  $\hat{a}_i^u$  and  $\hat{d}_i^u$  are the predicted aesthetic attributes and score distribution of the  $i$ -th image. Then, we leveraged the user's subjective attributes  $s_u$  and predicted objective attributes  $\hat{a}_i^u$  for interaction and fused the interactive feature  $I_i^u$  and score distribution  $\hat{d}_i^u$  to obtain a personalized aesthetic score, which can be computed by

$$\hat{y}_i^u = FC_{\theta_s}(\hat{a}_i^u) + MLP_{\theta_r}(I_i^u), \quad (9)$$

where  $\hat{y}_i^u$  indicates the predicted aesthetic score. In general, a specific user can only provide a small number of annotated samples for model fine-tuning. Therefore, we only optimized the parameters of the MLP and FC layers ( $\theta_r$  and  $\theta_s$ ) by using the  $l_2$  loss function, which is formulated as

$$\mathcal{L}_u = \frac{1}{N_u} \sum_{i=1}^{N_u} (y_i^u - \hat{y}_i^u)^2. \quad (10)$$

In this manner, fine-tuning a small number of parameters ( $\theta_r$  and  $\theta_s$ ) with annotated samples can enable the prior model to be easily transferred to the PIAA model of the specific user. For a testing image, we fed it into the PIAA model and obtained the user's personalized aesthetic score for the image.

## 4. Experimental Results

In this section, we employ extensive experiments to verify the effectiveness of our PIAA-MIR, which were mainly performed on three public PIAA databases: PAPA (<https://web.xidian.edu.cn/ldli/en/dataset.html>, accessed on 7 October 2022) [10], FLICKR-AES, and REAL-CUR (<https://github.com/alanspike/personalizedImageAesthetics>, accessed on 7 October 2022) [22].

### 4.1. Databases

The PAPA [10] database contains 31,220 images with rich annotation rated by 438 users. Besides the aesthetic score, each image was annotated by several users with seven objective attributes: *composition*, *light*, *color*, *depth of field*, *object emphasis*, *content*, and *scene category*. For each user, the database also provided some subjective attributes: *age*, *gender*, *education experience*, *artistic experience*, *photographic experience*, and *Big-Five personality traits* [41]. The education experience was divided into six steps: junior high school, senior high school, technical secondary school, junior college, and university. The artistic experience and photographic experience included beginner, competent, proficient, and expert. The Big-Five traits (extroversion (E), agreeableness (A), neuroticism (N), openness (O), and conscientiousness (C)) of each user were collected by asking them to fill in the BFI-10 questionnaire [42]. In this database, 40 users and their corresponding rated images were randomly selected as the testing set, and the remaining users and their corresponding rated images were used as the training set. Therefore, we can use the interaction between the multiple attributes of images and users to train the proposed PIAA-MIR model.

The FLICKR-AES [22] database contains 40,000 images rated by 210 users. Among them, 173 users and their rated 35,263 images were chosen as the training set, and the remaining 37 users and their rated 4,737 images were used as the testing set. Since the



database only provided each user's personalized aesthetic score for the images, we could use the general aesthetic extraction module trained by the PAPA database to obtain multiple attributes of images. Similar to [27], we also could leverage users' aesthetic ratings on images to obtain their Big-Five personality traits. In this way, the Big-Five personality traits were used as subjective attributes to interact with the objective attributes of images to train our PIAA-MIR model.

The **REAL-CUR** [22] is a relatively small database that contains 14 users and their personalized aesthetic ratings on images in their own photo albums. Each photo album only consists of images ranging from 197 to 222. Due to the small number of users in this database, we directly fine-tuned the prior model trained on the PAPA database with the PIAA tasks of these 14 users, which can verify the generalization performance of the proposed prior model for inferring users' personalized aesthetics in a real scenario.

In the following experiments on these three databases, all aesthetic scores and numerical attributes were normalized to the range of 0 to 1, and higher values indicate stronger aesthetics and attributes.

#### 4.2. Experimental Settings

**Implementation details:** The initialized parameters of our CNN model ( $f_\theta$ ) came from ResNet50 [38], which is pre-trained on ImageNet [39]. In the multi-attribute interaction reasoning network,  $MLP_{\theta_r}$  consists of two FC layers with 1024 nodes and 1 node. All parameters of FC layers were randomly initialized. In the training process of the generic aesthetics extraction module and multi-attribute interaction reasoning network, we set the initial learning rate to  $5 \times 10^{-5}$ , and the learning rate was multiplied by 0.1 every 5 epochs. Besides, the batch size and the number of epochs were set to 100 and 20, respectively. In the PIAA model's fine-tuning, the number of epochs was set to 5, and the learning rate was set to  $1 \times 10^{-5}$ . The proposed PIAA-MIR was performed on PyTorch, and Adam was used as the optimizer of our model.

**Evaluation criterion:** As with the previous approaches [27,28], the Spearman rank-order correlation coefficient (SROCC) was adopted to evaluate the effectiveness of the PIAA models in predicting users' personalized aesthetic scores on images. The values of the SROCC range from  $-1$  to  $1$ , and higher values of the SROCC indicate better performance of the PIAA methods.

#### 4.3. Comparing with the State-of-the-Art PIAA Methods

Since PAPA is a recently released PIAA database, only a few results of the PIAA models have been reported in this database [10]. To further examine the performance of the proposed method, we also compared our PIAA-MIR with a generic aesthetic prior-based method (PA\_IAA [8]) and two personalized aesthetic prior-based methods (BLG-PIAA [28] and PIAA-SOA [27]). Similar to [10], we randomly selected 40 users for testing and report the average results of 10 repeated experiments. For each user, 10 or 100 images rated by the user were selected to fine-tune the prior model for obtaining the PIAA model. To avoid random bias, the fine-tuning process for each user was repeated 10 times, and the average results and the corresponding standard deviation are reported.

Table 1 lists the comparison results of our PIAA-MIR with several PIAA methods on the PAPA database [10], where the mean SROCC results of 40 testing users were used as the final results, and the best results are highlighted in bold font. Overall, our PIAA-MIR method achieved the best performance when fine-tuning with 10 or 100 images, which indicates the effectiveness of the proposed multi-attribute interaction-based prior model. Compared with the PIAA models only using a generic aesthetic prior (PAPA (unconditional) and PA\_IAA) or a personalized aesthetic prior (BLG-PIAA and PIAA-SOA), PIAA-MIR achieved superior performance, demonstrating that it is efficient in jointly learning the prior model from the perspectives of both users and images. In addition, the proposed PIAA-MIR outperformed the three types of conditional PIAA models (PAPA (artistic), PAPA (photographic), and PAPA (photographic)), which shows that it is more effective at

learning users' aesthetic preferences through multiple attributes' interaction than directly embedding subjective attributes.

**Table 1.** SROCC results of our PIAA-MIR with several PIAA methods on the PAPA database [10]. PAPA (unconditional) indicates the unconditional PIAA model proposed by the authors of PAPA. Similarly, PAPA (artistic), PAPA (photographic), and PAPA (photographic) denote the PIAA model by embedding three types of conditional information (artistic experience, photographic experience, and personality traits).

Methods	10 Images	100 Images
PAPA (unconditional) [10]	0.681 ± 0.0015	0.695 ± 0.0014
PAPA (artistic) [10]	0.686 ± 0.0016	0.698 ± 0.0012
PAPA (photographic) [10]	0.683 ± 0.0014	0.698 ± 0.0010
PAPA (personality) [10]	0.691 ± 0.0009	0.705 ± 0.0015
PA_IAA [8]	0.683 ± 0.0013	0.690 ± 0.0016
BLG-PIAA [28]	0.688 ± 0.0015	0.697 ± 0.0013
PIAA-SOA [27]	0.692 ± 0.0014	0.703 ± 0.0012
<b>PIAA-MIR</b>	<b>0.702 ± 0.00010</b>	<b>0.716 ± 0.0008</b>

As with the experimental setup in [27], we verified the performance of the proposed method on the FLICKR-AES and REAL-CUR databases. In Table 2, we summarize the average SROCC results of the proposed PIAA-MIR and several state-of-the-art methods on the 37 testing users of the FLICKR-AES database, where the best results are highlighted in bold font. From the table, we can see that the proposed method significantly outperformed all the PIAA methods, except PIAA-SOA. This illustrates that the objective attributes of images learned on the PAPA database are also beneficial to building the prior model on the FLICKR-AES database. Compared with PIAA-SOA, which directly integrates objective attributes and subjective attributes, the proposed method utilizes multi-attribute interaction to learn better personalized aesthetic prior knowledge for individual users. To verify the effectiveness of the proposed prior model in adapting to users' personalized aesthetic preferences in real scenarios, we list the average SROCC results of our PIAA-MIR and three PIAA methods reported in [27] on the 14 album users of the REAL-CUR database in Table 3. As shown in the table, the proposed PIAA-MIR yielded the best performance in learning the aesthetic preferences of individual users in real applications. This further proves that our prior model learned on the PAPA database also has satisfactory generalization performance for users of other databases.

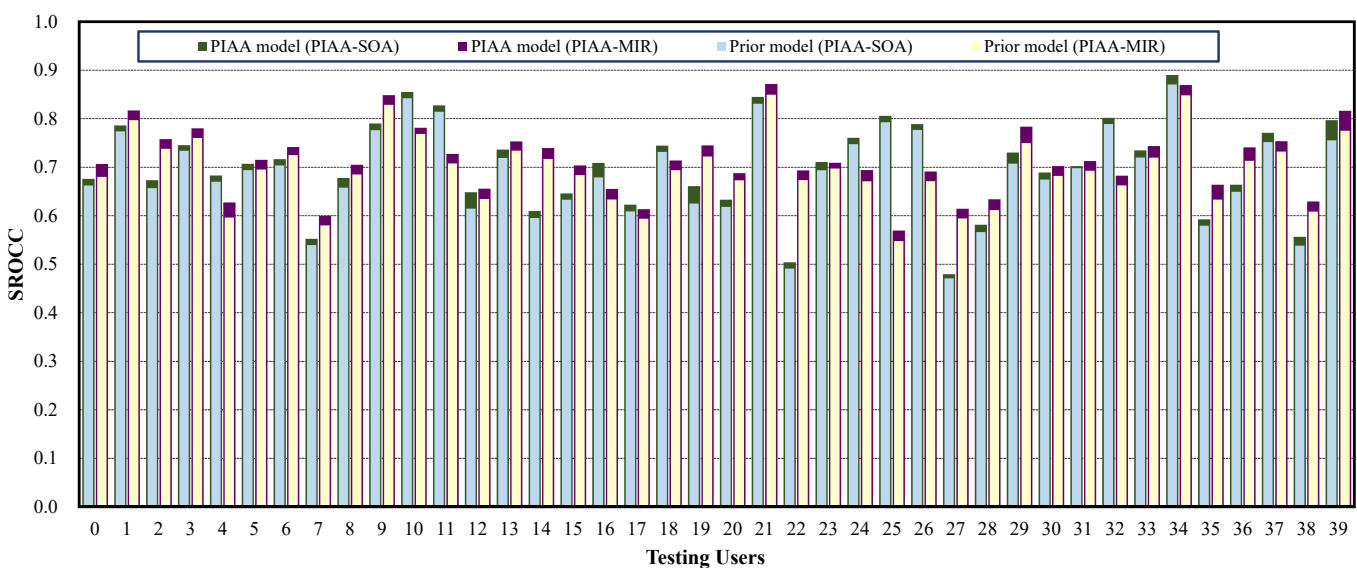
**Table 2.** SROCC results of our PIAA-MIR with several PIAA methods on the FLICKR-AES database [22].

Methods	10 Images	100 Images
PAM (attribute) [22]	0.518 ± 0.003	0.539 ± 0.013
PAM (content) [22]	0.515 ± 0.004	0.535 ± 0.017
PAM (content and attribute) [22]	0.520 ± 0.003	0.553 ± 0.012
USAR_PPR [23]	0.521 ± 0.002	0.544 ± 0.007
USAR_PAD [23]	0.520 ± 0.003	0.537 ± 0.003
USAR_PPR&PAD [23]	0.525 ± 0.004	0.552 ± 0.015
ML-PIAA [25]	0.522 ± 0.005	0.562 ± 0.015
PA_IAA [8]	0.543 ± 0.003	0.639 ± 0.011
BLG-PIAA [28]	0.561 ± 0.005	0.669 ± 0.013
UG-PIAA [26]	0.559 ± 0.002	0.660 ± 0.013
PIAA-SOA [27]	0.618 ± 0.006	0.691 ± 0.015
<b>PIAA-MIR</b>	<b>0.621 ± 0.005</b>	<b>0.713 ± 0.00016</b>

**Table 3.** SROCC results of our PIAA-MIR with three PIAA methods on the REAL-CUR database [22].

Methods	10 Images	100 Images
PA_IAA [8]	$0.443 \pm 0.004$	$0.562 \pm 0.013$
BLG-PIAA [28]	$0.448 \pm 0.007$	$0.578 \pm 0.015$
PIAA-SOA [27]	$0.487 \pm 0.006$	$0.589 \pm 0.014$
<b>PIAA-MIR</b>	<b><math>0.498 \pm 0.008</math></b>	<b><math>0.606 \pm 0.013</math></b>

To further verify the efficiency of our method in learning each user’s personalized aesthetic experience from the proposed prior model, we examined the performance of the prior model and the PIAA model fine-tuned on 100 images of each testing user from the PAPA database [10]. To highlight the comparative results, we compared our PIAA-MIR with the state-of-the-art PIAA-SOA and show the average SROCC results of 10 experiments on 40 testing users in Figure 3. For both PIAA-SOA and the proposed PIAA-MIR, the PIAA model yielded better performance than the prior model. For most users (27 out of 40), PIAA-MIR outperformed PIAA-SOA in terms of the prior model (0.695 versus 0.686), which shows the effectiveness of the proposed prior model in capturing the personalized aesthetic experiences of individual users by using the interaction between multiple objective and subjective attributes. In addition, when the prior model was fine-tuned on 100 images rated by individual users, our method was also superior to PIAA-SOA in transferring users’ personalized aesthetics from the prior model (0.021 (from 0.695 to 0.716) versus 0.018 (from 0.686 to 0.703)). In summary, the proposed PIAA-MIR builds a robust prior model through multi-attribute interaction, which can easily adapt to personalized aesthetic preferences with a small number of annotated samples.



**Figure 3.** SROCC results of PIAA-SOA [27] and our PIAA-MIR on the 40 testing users of the PAPA database [10]. The testing results of the prior model and PIAA model on each user are shown. Specifically, the testing results of PIAA-SOA are displayed with blue and green bars, and the testing results of PIAA-MIR are displayed with yellow and purple bars.

#### 4.4. Ablation Study

To further examine the contribution of each module in the proposed multi-attribute interactive reasoning network for learning users’ personalized aesthetic preferences for images, an ablation study was conducted on the PAPA database [10]. In the generic aesthetics extraction module, we removed the prediction branch of multiple objective attributes and only leveraged multiple subjective attributes and score distributions to predict personalized aesthetic scores, which is termed “PIAA-MIR w/o objective”. In the multi-attribute interactive reasoning network, we replaced multiple subjective attributes with a one-hot encoding

vector to characterize users (PIAA-MIR w/o subjective). We replaced the multi-attribute interaction with simple attributes combined to predict personalized aesthetic scores, which is called “PIAA-MIR w/o interaction”. In addition, to compare the prior model learned only from generic aesthetics or personalized aesthetics, we introduced the baseline model by only training the generic aesthetics extraction module from the perspective of images (Baseline (generic)) or the multi-attribute interaction reasoning network from the perspective of users (Baseline (personalized)).

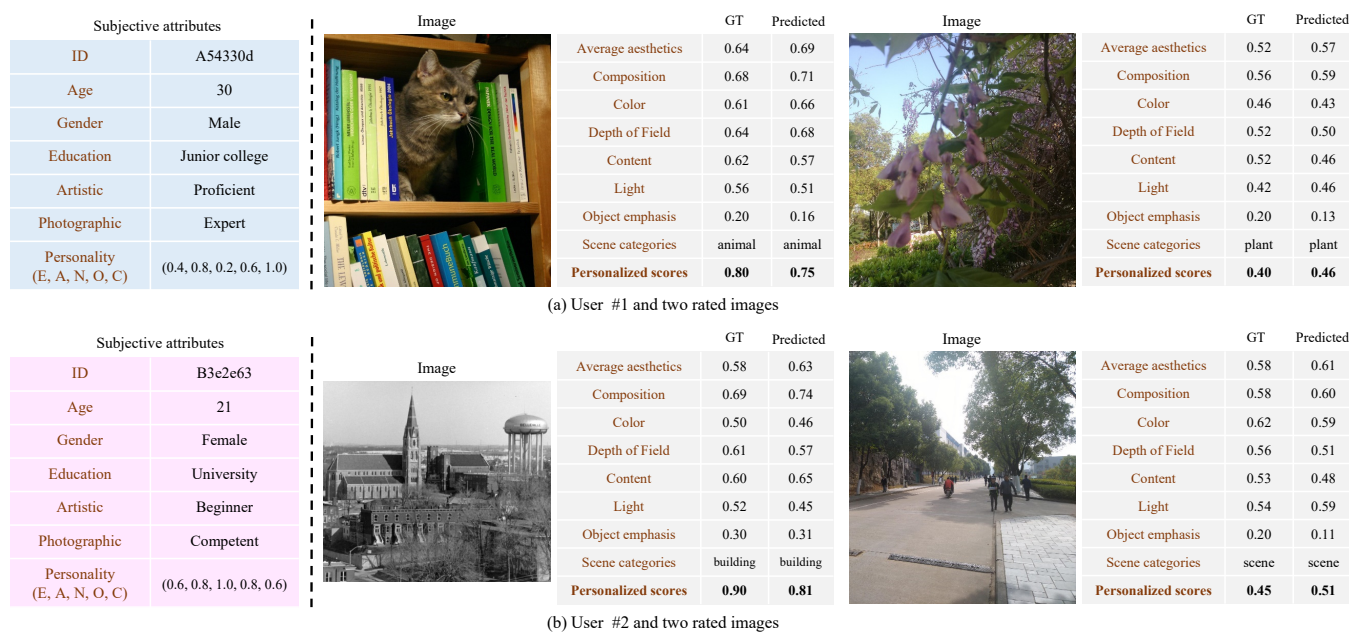
Table 4 lists the test results of the ablation experiments. As shown in the table, the full version of PIAA-MIR obtained the best results on the testing users of the PAPA database. Compared with the baseline model learned only from generic aesthetics (Baseline (generic)) or personalized aesthetics (Baseline (personalized)), the proposed PIAA-MIR yielded significant performance improvements, which shows that it is efficient at learning a prior model from the perspectives of both images and users. When eliminating multiple objective attributes (PIAA-MIR w/o objective) or multiple subjective attributes (PIAA-MIR w/o subjective) in our model, PIAA-MIR showed worse prediction performance in learning personalized aesthetics, which demonstrates the importance of embedding subjective and objective attributes in the proposed PIAA-MIR. Besides, PIAA-MIR was also superior to “PIAA-MIR w/o interaction”, which indicates that the proposed multi-attribute interaction is crucial for exploring the underlying factors for users’ personalized aesthetic experiences. All in all, the above modules contributed to promoting the evaluation performance of the proposed method.

**Table 4.** SROCC results of our PIAA-MIR on the PAPA database [10] by eliminating different ablation modules, where the best results of fine-tuning on 10 and 100 images are shown in boldface.

Methods	10 Images	100 Images
Baseline (generic)	0.679 ± 0.0014	0.692 ± 0.0015
Baseline (personalized)	0.682 ± 0.0015	0.698 ± 0.0016
PIAA-MIR w/o objective	0.689 ± 0.0011	0.700 ± 0.0011
PIAA-MIR w/o subjective	0.684 ± 0.0013	0.693 ± 0.0014
PIAA-MIR w/o interaction	0.696 ± 0.0012	0.707 ± 0.0010
<b>PIAA-MIR</b>	<b>0.702 ± 0.00010</b>	<b>0.716 ± 0.0008</b>

#### 4.5. Visual Analysis

To intuitively show how PIAA-MIR leverages the interaction between multiple subjective and objective attributes for personalized aesthetic preferences reasoning, we randomly selected two testing users, as well as two testing images rated by them from the PAPA database [10]. The predicted results of our method are shown in Figure 4. We can see from the figure that the predicted attributes and aesthetic scores of the proposed PIAA-MIR for the four images were close to the ground truth (GT) results, which indicates that the proposed generic aesthetics extraction module is efficient in predicting aesthetic attributes and the score distribution. Since User #1 is a man with expert photography experience, he tends to give higher aesthetic ratings to images with better composition and content. In addition, User #1 is also a person with strong agreeableness and conscientiousness, so he prefers the left image containing animals. By contrast, User #2 is a person with strong neuroticism and has preliminary art and photography experience. Although the two images rated by User #2 have the same average aesthetics, the user’s personalized aesthetic scores for these two images differ greatly. This is because neurotic people prefer images with dim light and monotonous color [41], which led User #2 to give a higher aesthetic score to the left image than the right image. From the above analysis, we can draw a conclusion that the multiple objective attributes of images and the multiple subjective attributes of users jointly affect users’ personalized aesthetic experiences for images, and the proposed multi-attribute interaction can effectively reveal the potential impact relationship between them.



**Figure 4.** Qualitative results of the proposed model on two testing users from the PAPA database [10]. The identification (ID) information and some subjective attributes of these two users are shown on the left side. The average aesthetics of score distribution, objective attributes, and personalized scores of images are shown on the right side. For comparison, we show both the ground truth (GT) and predicted results, where aesthetic scores and numerical attributes are normalized to the range of 0 to 1 and higher values indicate stronger aesthetics and attributes.

### 5. Conclusions

In this paper, we introduced a personalized image aesthetics assessment method via multi-attribute interactive reasoning (PIAA-MIR). Compared with existing PIAA methods, the proposed method can effectively reason users’ personalized image aesthetic experiences, which benefits from learning the prior model for PIAA from the perspectives of both images and users. Specifically, the proposed generic aesthetics extraction module showed its efficiency in predicting multiple aesthetic attributes and score distributions of images. In addition, the multi-attribute interaction-based prior model learned from extensive users’ PIAA tasks can capture the robust impact of multiple subjective and objective attributes on users’ personalized aesthetic preferences for images. Therefore, when an individual user only provides a small number of annotation samples, the proposed multi-attribute interaction can use this robust interactive relationship to effectively transfer the prior model to the PIAA model for the individual user. The experimental results and visual analysis of three PIAA databases demonstrated that the proposed PIAA model is effective in reasoning individual users’ personalized visual aesthetics. In the future, our method will highlight a novel strategy to analyze the implicit reasons for personalized aesthetic preferences from the perspectives of both users and images.

**Author Contributions:** Conceptualization, H.Z. and Y.Z.; methodology, H.Z. and Z.S.; software, H.Z.; validation, W.D., G.W. and Q.L.; formal analysis, W.D.; investigation, H.Z.; resources, H.Z.; data collection, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, Y.Z. and Z.S.; visualization, H.Z.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (No. 62101555), the Natural Science Foundation of Jiangsu Province (No. BK20210488), the Project funded by China Postdoctoral Science Foundation (No. 2022M713379), and the Fundamental Research Funds for the Central Universities (No. 2021QN1071). It was also partially supported by the National Natural Science Foundation of China (Nos. 62272461, 62106268, and 62002360), the Natural Science Foundation of Jiangsu Province (No. BK20201346), the High-Level Talent Program for Innovation



and Entrepreneurship (ShuangChuang Doctor) of Jiangsu Province (No. JSSCBS20211220), and the Six Talent Peaks High-level Talents in Jiangsu Province (No. 2015-DZXX-010).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The experiment used three public PIAA databases, which are available in the PAPA [10], FLICKR-AES [22], and REAL-CUR [22] databases.

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

## References

1. Deng, Y.; Chen, C.L.; Tang, X. Image Aesthetic Assessment: An Experimental Survey. *IEEE Signal Process. Mag.* **2017**, *34*, 80–106. [[CrossRef](#)]
2. Talebi, H.; Milanfar, P. NIMA: Neural Image Assessment. *IEEE Trans. Image Process.* **2018**, *27*, 3998–4011. [[CrossRef](#)]
3. Ma, W.; Qin, J.; Xiang, X.; Tan, Y.; He, Z. Searchable Encrypted Image Retrieval Based on Multi-Feature Adaptive Late-Fusion. *Mathematics* **2020**, *8*, 1019. [[CrossRef](#)]
4. Karlsson, K.; Jiang, W.; Zhang, D.Q. Mobile photo album management with multiscale timeline. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 1061–1064. [[CrossRef](#)]
5. Lozano-Vázquez, L.V.; Miura, J.; Rosales-Silva, A.J.; Luviano-Juárez, A.; Mújica-Vargas, D. Analysis of Different Image Enhancement and Feature Extraction Methods. *Mathematics* **2022**, *10*, 2407. [[CrossRef](#)]
6. Esser, P.; Rombach, R.; Ommer, B. Taming Transformers for High-Resolution Image Synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12873–12883. [[CrossRef](#)]
7. Zhang, Y.; Yamasaki, T. Style-Aware Image Recommendation for Social Media Marketing. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual, China, 20–24 October 2021; pp. 3106–3114. [[CrossRef](#)]
8. Li, L.; Zhu, H.; Zhao, S.; Ding, G.; Lin, W. Personality-Assisted Multi-Task Learning for Generic and Personalized Image Aesthetics Assessment. *IEEE Trans. Image Process.* **2020**, *29*, 3898–3910. [[CrossRef](#)]
9. Zhang, X.; Zhang, X.; Xiao, Y.; Liu, G. Theme-Aware Semi-Supervised Image Aesthetic Quality Assessment. *Mathematics* **2022**, *10*, 2609. [[CrossRef](#)]
10. Yang, Y.; Xu, L.; Li, L.; Qie, N.; Li, Y.; Zhang, P.; Guo, Y. Personalized Image Aesthetics Assessment With Rich Attributes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 19861–19869. [[CrossRef](#)]
11. Murray, N.; Marchesotti, L.; Perronnin, F. AVA: A large-scale database for aesthetic visual analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2408–2415. [[CrossRef](#)]
12. Kao, Y.; He, R.; Huang, K. Deep Aesthetic Quality Assessment With Semantic Information. *IEEE Trans. Image Process.* **2017**, *26*, 1482–1495. [[CrossRef](#)]
13. Kong, S.; Shen, X.; Lin, Z.; Mech, R.; Fowlkes, C. Photo Aesthetics Ranking Network with Attributes and Content Adaptation. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 662–679. [[CrossRef](#)]
14. Pan, B.; Wang, S.; Jiang, Q. Image Aesthetic Assessment Assisted by Attributes through Adversarial Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January 27–1 February 2019; pp. 679–686. [[CrossRef](#)]
15. Zhang, X.; Gao, X.; Lu, W.; He, L. A Gated Peripheral-Foveal Convolutional Neural Network for Unified Image Aesthetic Prediction. *IEEE Trans. Multimedia* **2019**, *21*, 2815–2826. [[CrossRef](#)]
16. Zeng, H.; Cao, Z.; Zhang, L.; Bovik, A.C. A Unified Probabilistic Formulation of Image Aesthetic Assessment. *IEEE Trans. Image Process.* **2020**, *29*, 1548–1561. [[CrossRef](#)]
17. She, D.; Lai, Y.K.; Yi, G.; Xu, K. Hierarchical Layout-Aware Graph Convolutional Network for Unified Aesthetics Assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 8475–8484. [[CrossRef](#)]
18. Zhang, X.; Song, Q.; Liu, G. Multimodal Image Aesthetic Prediction with Missing Modality. *Mathematics* **2022**, *10*, 2312. [[CrossRef](#)]
19. Zhang, J.; Yang, Y.; Zhuo, L.; Tian, Q.; Liang, X. Personalized Recommendation of Social Images by Constructing a User Interest Tree with Deep Features and Tag Trees. *IEEE Trans. Multimedia* **2019**, *21*, 2762–2775. [[CrossRef](#)]
20. Park, C.C.; Kim, B.; KIM, G. Towards Personalized Image Captioning via Multimodal Memory Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 999–1012. [[CrossRef](#)] [[PubMed](#)]
21. Bianco, S.; Cusano, C.; Piccoli, F.; Schettini, R. Personalized Image Enhancement Using Neural Spline Color Transforms. *IEEE Trans. Image Process.* **2020**, *29*, 6223–6236. [[CrossRef](#)] [[PubMed](#)]
22. Ren, J.; Shen, X.; Lin, Z.; Mech, R.; Foran, D.J. Personalized Image Aesthetics. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 638–647. [[CrossRef](#)]
23. Lv, P.; Wang, M.; Xu, Y.; Peng, Z.; Sun, J.; Su, S.; Zhou, B.; Xu, M. USAR: An Interactive User-Specific Aesthetic Ranking Framework for Images. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 1328–1336. [[CrossRef](#)]

24. Wang, G.; Yan, J.; Qin, Z. Collaborative and Attentive Learning for Personalized Image Aesthetic Assessment. In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 957–963. [\[CrossRef\]](#)
25. Wang, W.; Su, J.; Li, L.; Xu, X.; Luo, J. Meta-Learning Perspective for Personalized Image Aesthetics Assessment. In Proceedings of the IEEE International Conference on Image Processing, Taipei, Taiwan, 22–25 September 2019; pp. 1875–1879. [\[CrossRef\]](#)
26. Lv, P.; Fan, J.; Nie, X.; Dong, W.; Jiang, X.; Zhou, B.; Xu, M.; Xu, C. User-Guided Personalized Image Aesthetic Assessment based on Deep Reinforcement Learning. *IEEE Trans. Multimedia* **2021**, 1–14. [\[CrossRef\]](#)
27. Zhu, H.; Zhou, Y.; Li, L.; Li, Y.; Guo, Y. Learning Personalized Image Aesthetics from Subjective and Objective Attributes. *IEEE Trans. Multimedia* **2021**, 1–12. [\[CrossRef\]](#)
28. Zhu, H.; Li, L.; Wu, J.; Zhao, S.; Ding, G.; Shi, G. Personalized Image Aesthetics Assessment via Meta-Learning with Bilevel Gradient Optimization. *IEEE Trans. Cybern.* **2022**, *52*, 1798–1811. [\[CrossRef\]](#)
29. Hou, J.; Lin, W.; Yue, G.; Liu, W.; Zhao, B. Interaction-Matrix Based Personalized Image Aesthetics Assessment. *IEEE Trans. Multimedia* **2022**, 1–16. [\[CrossRef\]](#)
30. Kucer, M.; Loui, A.C.; Messenger, D.W. Leveraging Expert Feature Knowledge for Predicting Image Aesthetics. *IEEE Trans. Image Process.* **2018**, *27*, 5100–5112. [\[CrossRef\]](#)
31. Vinciarelli, A.; Mohammadi, G. A Survey of Personality Computing. *IEEE Trans. Affect. Comput.* **2014**, *5*, 273–291. [\[CrossRef\]](#)
32. He, X.; Du, X.; Wang, X.; Tian, F.; Tang, J.; Chua, T. Outer Product-based Neural Collaborative Filtering. In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 2227–2233. [\[CrossRef\]](#)
33. Zeki, S. Clive Bell’s “Significant Form” and the neurobiology of aesthetics. *Front. Hum. Neurosci.* **2013**, *7*, 730. [\[CrossRef\]](#)
34. Perona, F.R.; Gallego, M.J.F.; Callejón, J.M.P. An Application for Aesthetic Quality Assessment in Photography with Interpretability Features. *Entropy* **2021**, *23*, 1389. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Gelli, F.; Uricchio, T.; He, X.; Del Bimbo, A.; Chua, T.S. Learning Subjective Attributes of Images from Auxiliary Sources. In Proceedings of the ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2263–2271. [\[CrossRef\]](#)
36. Kim, W.H.; Choi, J.H.; Lee, J.S. Objectivity and Subjectivity in Aesthetic Quality Assessment of Digital Photographs. *IEEE Trans. Affect. Comput.* **2020**, *11*, 493–506. [\[CrossRef\]](#)
37. Zhao, W.; Wang, B.; Ye, J.; Yang, M.; Zhao, Z.; Luo, R.; Qiao, Y. A Multi-task Learning Approach for Image Captioning. In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 1205–1211. [\[CrossRef\]](#)
38. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)
39. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
40. Levina, E.; Bickel, P. The Earth Mover’s distance is the Mallows distance: Some insights from statistics. In Proceedings of the IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 251–256. [\[CrossRef\]](#)
41. Zhu, H.; Li, L.; Zhao, S.; Jiang, H. Evaluating attributed personality traits from scene perception probability. *Pattern Recognit. Lett.* **2018**, *116*, 121–126. [\[CrossRef\]](#)
42. Rammstedt, B.; John, O.P. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *J. Res. Pers.* **2007**, *41*, 203–212. [\[CrossRef\]](#)