This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

Learning Personalized Image Aesthetics from Subjective and Objective Attributes

Hancheng Zhu, Yong Zhou, Leida Li, *Member, IEEE,* Yaqian Li, and Yandong Guo

***Abstract*—Due to the widespread popularity of social media, researchers have developed a strong interest in learning the personalized image aesthetics of online users. Personalized image aesthetics assessment (PIAA) aims to study the aesthetic preferences of individual users for images, which should be affected by the properties of both users and images. Existing PIAA approaches usually use the generic aesthetics learned from images as a prior model and adapt it to PIAA models through a small number of data annotated by individual users. However, the prior model merely learns the objective attributes of images, which is agnostic to the subjective attributes of users, complicating efficient learning of the personalized image aesthetics of individual users. Therefore, we propose a personalized image aesthetics assessment method that integrates the subjective attributes of users and objective attributes of images simultaneously. To characterize these two attributes jointly, an attribute extraction module is introduced to learn users' personality traits and image aesthetic attributes. Then, an aesthetic prior model is built from numerous individual users' annotated data, which leverages the personality traits of users and the aesthetic attributes of rated images as prior knowledge to model both the image aesthetic distribution and users' residual scores relative to generic aesthetics simultaneously. Finally, a PIAA model is obtained by fine-tuning the aesthetic prior model with an individual user's annotated data. Experiments demonstrate that the proposed method is superior to existing PIAA methods in learning individual users' personalized image aesthetics.**

***Index Terms*—Personalized image aesthetics assessment, aesthetic attributes, personality traits, aesthetic prior model, aesthetic preferences**

I. Introduction

In the past decade, the prevalence of mobile devices has prompted an increasing number of people to consume, entertain and learn on social media. To engage potential users, many social media platforms (e.g. Facebook, Flickr, and Instagram) usually leverage beautiful images to meet their visual demands. Consequently, the advancement of social media is a fundamental issue in understanding users' aesthetic perception of images and automatically assess image aesthetics perceived by users. Recently, image aesthetics assessment (IAA) has received extensive attention in computer vision and multimedia communities [1] and is extremely valuable in promoting the development of many applications, e.g. image enhancement [2], [3], photo recommendation [4], product retrieval [5], album management [6] and UI design [7].

IAA is a challenging task in that people's aesthetic preferences for images are inherently subjective, which complicates effective quantification [8]. Previous IAA approaches mainly focus on most people's aesthetic experience of images and primarily employ general photography rules [9] for generic IAA (GIAA) models, which can convert the image aesthetics perceived by most people into two categories (high-quality or low-quality) [10]–[14] or a quality score [15]–[17]. However, since different people usually have diversified tastes for images, early GIAA models cannot thoroughly describe the image aesthetics of people with different subjectivity [18]. Therefore, recent GIAA models mainly focus on aesthetic distribution prediction [19]–[22], which can infer the aesthetic distribution of images rated by multiple users. While the aesthetic distribution of images can reveal users' subjectivity in visual aesthetics to a certain extent, it only characterizes their aesthetic experience from the perspective of images, which cannot be directly applied to "user-centred" application scenarios, such as personalized consumption [23] and personalized image enhancement [24]. Thus, very recently, a few personalized IAA (PIAA) models have been proposed for learning the personalized image aesthetics of individual users [25]–[30].

PIAA learns an individual user's aesthetic evaluation of images, that is, a PIAA model needs to be built for each user [25]. Hence, the PIAA task is to train a PIAA model through a set of images annotated by a user. In the PIAA task, the greatest challenge is that a user usually annotates only a small number of images, which is insufficient to directly train an efficient PIAA model based on a deep network [30]. Therefore, existing PIAA methods mainly take the generic image aesthetics rated by a large number of users as a prior model and utilize a small amount of annotated data to fine-tune it for modeling individual users' personalized image aesthetics [26]–[30]. However, the prior model is learned from the objective attributes extracted from images and is unable to acquire users' subjective attributes. The subjective attributes of users (e.g., age, gender, emotion, and personality) are the inherent factors that determine their unique aesthetic experience for images [18], [31], which can influence their different aesthetic

This work was supported by the National Natural Science Foundation of China under Grants 62101555, 62171340, 61771473 and 61991451, the Natural Science Foundation of Jiangsu Province under Grants BK20210488, BK20201346 and BK20181354, the OPPO Research Fund, the Key Project of Shaanxi Provincial Department of Education (Collaborative Innovation Center) under Grant 20JY024, the Six Talent Peaks High-level Talents in Jiangsu Province under Grants 2015-DZXX-010 and XYDXX-063, and the Fundamental Research Funds for the Central Universities under Grants 2021QN1071 and JBF211902. (*Corresponding authors: Yong Zhou; Leida Li.*)

H. Zhu and Y. Zhou are with the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China (e-mails: zhuhancheng@cumt.edu.cn; yzhou@cumt.edu.cn).

L. Li is with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: ldli@xidian.edu.cn).

Y. Li and Y. Guo are with the OPPO Research Institute, Shanghai 200032, China (e-mails: liyaqian@oppo.com; guoyandong@oppo.com).

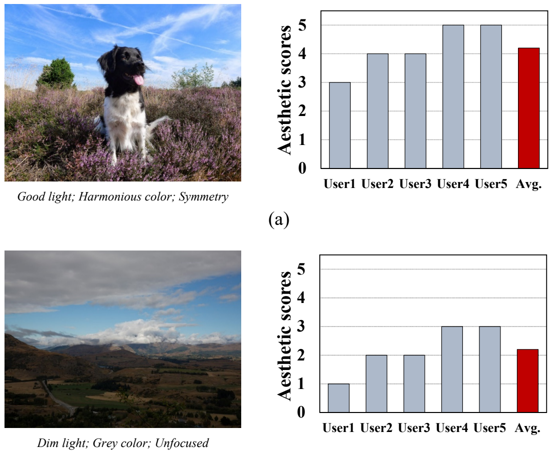


Fig. 1. Two example images and relevant attributes from the FLICKR-AES database [25], together with the aesthetic scores and corresponding average scores rated by five users. The range of scores is between 1 and 5. The higher the scores, the higher the aesthetics of images.

preferences for the objective attributes of images (e.g. light, color, and composition) [25]. Therefore, transferring the prior model to an effective PIAA model without considering the subjective attributes of users is difficult.

To address the above issue, one possible solution is to jointly characterize users' aesthetic preferences for images from both the subjective attributes of users [31] and the objective attributes of images [15] when learning the aesthetic prior model for PIAA. For example, Fig. 1 shows two example images and some relevant attributes from the FLICKR-AES database [25], together with the aesthetic scores and corresponding average scores rated by five users. The scores range from 1 to 5, with higher scores corresponding to higher image aesthetics. The figure shows that the average score of Fig. 1(a) is higher than that of Fig. 1(b). The underlying reason is that Fig. 1(a) has better objective attributes than Fig. 1(b), which causes Fig. 1(a) to receive higher aesthetic ratings from most users. In addition, the aesthetic ratings of different users for the same image are quite inconsistent, possibly because the aesthetic preferences of individual users are also affected by their own subjective attributes [31]. Therefore, the objective attributes of images are common aesthetic rules, which can be used to measure the generic image aesthetics of most people [15], and the subjective attributes of users are the critical factors that determine the difference between their personalized aesthetic preferences and generic image aesthetics [31].

Inspired by the above facts, we propose a personalized image aesthetics assessment approach that integrates subjective and objective attributes from the perspective of both users and images. To obtain the subjective attributes of users, it is necessary to make use of the stable psychological characteristics that influence people's visual aesthetic preferences. Recently, some studies have shown that personality traits [32] can be used to stably quantify the subjective attributes of users in their visual aesthetic preferences [31], [33], [34]. That is to say, users' personality traits are closely related to their aesthetic preferences for images. For example, an extrovert tends to assign high aesthetic ratings to images with outdoor scenes, while an open person prefers artistic pictures [35]. Consequently, personality traits can be adopted to measure users' subjective attributes in our work. Moreover, the objective attributes of images can be measured by aesthetic attributes, such as the rule of thirds, depth of field, and color harmony [15], [16].

To leverage personality traits and aesthetic attributes to characterize users and images, respectively, an attribute extraction module is introduced for modeling the personality traits of users from a set of liked images and learning the aesthetic attributes of images simultaneously. Then, we embed users' personality traits and the aesthetic attributes of the rated images into an aesthetic prior model from extensive users' PIAA tasks. In the aesthetic prior model, the generic aesthetics of images and the aesthetic differences among different users can be efficiently learned from the attributes of both users and images. For the PIAA task of a target user, the prior model would be easily adapted to a more efficient PIAA model by extracting the user's personality traits and image aesthetic attributes. In summary, the main contributions of the proposed method are threefold:

- We explore users' personalized aesthetic preferences for images from the perspective of both users and images. Based on the PIAA tasks of numerous users, an efficient aesthetic prior model is learned by embedding the subjective attributes of users and the objective attributes of rated images.
- We design an attribute extraction module that can predict users' personality traits and image aesthetic attributes simultaneously, where users' personality traits are inferred from a set of their rated images.
- We propose a PIAA method that can be easily adapted from our aesthetic prior model. Experimental results on several PIAA databases demonstrate that the proposed method significantly outperforms the state-of-the-art PIAA models.

The remainder of this paper is organized as follows. Section II provides a brief review of related works about PIAA methods. Section III presents the detail of the proposed PIAA model. The experimental results and visual analysis on several PIAA databases are given in Section IV. Finally, the conclusions are drawn in Section V.

II. Related Work

This section reviews some works related to our PIAA method that is achieved from the subjective attributes of users (personality traits) and the objective attributes of images (aesthetic attributes). Section II-A introduces some PIAA models, especially the PIAA models based on aesthetic attributes. In Section II-B, we briefly review several representative methods for image-based personality prediction.

A. Personalized Image Aesthetics Assessment

IAA utilizes a computational model to simulate human visual aesthetic experience and automatically assigns aesthetic scores to images [36]. To date, some efforts have achieved

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

remarkable progress on the GIAA model, which can be divided into three tasks: binary classification [10]–[13], score regression [15], [16] and distribution prediction [19]–[22]. Due to the aesthetic subjectivity of people [18], the above tasks only infer aesthetics from the perspective of images, which are difficult to meet the personalized aesthetic preferences of individual users [37]. Consequently, increasing attention has been given to the PIAA model for a specific user in recent years [25]–[30]. Since collecting a large number of images annotated by a specific user is difficult, existing methods usually use the trained GIAA model as prior knowledge and transfer this knowledge to obtain the PIAA model [38], which can be divided into three categories, i.e., user interaction-based approaches [26], [39], collaborative filtering-based approaches [27], [40], and aesthetic difference-based approaches [25], [28]–[30].

In the user interaction-based approach, Lv *et al.* [26] proposed a personalized image aesthetic ranking model, which extracted user-preferred images from a GIAA database through a retrieval method and used several aesthetic attributes of these images to characterize their visual aesthetic preferences. This method strictly relies on users' online interactions. In the collaborative filtering-based approach, Wang *et al.* [27] first designed a convolutional neural network (CNN) with user-image relationship coding to train the GIAA model. Then, a collaborative matrix was used to analyze the aesthetic correlation among different users and extract the PIAA model for a specific user. Nevertheless, one user may have a huge aesthetic difference from other users, which is not suitable to build an effective collaborative matrix.

Recently, the aesthetic difference-based PIAA approach has received increasing attention. Ren *et al.* [25] found that users' aesthetic preferences are closely related to image contents and aesthetic attributes. Based on the GIAA model, several image contents and aesthetic attributes were used to model the aesthetic difference between a user's personalized scores and the generic scores of images. Wang *et al.* [28] proposed a PIAA model via a meta-learning framework, which can learn the aesthetic difference between the generic score of an image and the personalized score of a specific user. Li *et al.* [29] proposed a method to model the aesthetic difference between image generic scores and users' personalized scores by using the personality traits of users who like images. Nevertheless, this approach is unable to acquire the personality traits of users in the PIAA task. Zhu *et al.* [30] leveraged deep meta-learning to integrate the aesthetic differences between different users from a large number of PIAA tasks into a prior model. However, all these approaches only characterize the aesthetic differences between various users by extracting the objective attributes in images, which is agnostic to the subjective attributes of users in their visual aesthetics. Hence, existing PIAA models cannot leverage users' subjective attributes to explicitly learn prior knowledge of their aesthetic preferences for images. With this problem, a feasible strategy is to capture both the subjective attributes of users and the objective attributes of images and fuse them into learning users' personalized image aesthetics.

B. Personality Prediction

In personality psychology, the Big-Five (BF) personality traits are one of the most influential models, which can be computed by users' behavioral cues externalized on social networks [41]. The BF personality traits are composed of five dimensions: *Openness* (tendency to be artistic, curious and imaginative), *Conscientiousness* (tendency to be responsible, organized and trustworthy), *Extraversion* (tendency to be positive, energetic, outgoing and talkative), *Agreeableness* (tendency to be appreciative, kind, generous and tolerant) and *Neuroticism* (tendency to be anxious, depressed, sensitive and unstable) [42]. In recent years, significant progress has been achieved in modeling users' BF personality traits [43] based on a set of their liked images [34], [35], [44], [45]. In [34], Segalin *et al.* proposed the PsychoFlickr database, which contains 60,000 liked images from 300 users on the Flickr website (200 images per user). Then, the BF personality traits of each user were obtained through the BIF-10 questionnaire and measured by five dimensions of openness, conscientiousness, extroversion, agreeableness, and neuroticism [43]. Earlier methods mainly used LASSO or SVM to model users' BF personality traits by extracting the low-level and high-level features of their preferred images [34], [44]. Recently, Zhu *et al.* [35] extracted the scene probability distribution of liked images based on a deep framework and used LASSO to predict users' personality traits. Recently, an end-to-end weakly supervised dual convolutional network has shown better performance by learning users' personality traits through a set of their liked images [45]. As shown by the above image-based personality prediction methods, users' aesthetic preferences are reflected in their liked images, which can be used to predict the BF personality traits that reveal their stable aesthetic subjectivity [31]. With this concept, we employ BF personality traits as the subjective attributes of users, which can be inferred by conveying their rated images into the personality prediction model.

III. Proposed Approach

This section details the proposed PIAA approach by fusing users' personality traits and image aesthetic attributes. In this way, the proposed PIAA model can effectively learn the personalized image aesthetics of users from the subjective attributes of users and the objective attributes of images, which is termed PIAA-SOA. An overview of our PIAA-SOA framework is shown in Fig. 2, which consists of three parts: an attribute extraction module, an aesthetic prior model, and a personalized aesthetics model. In the first part, users' BF personality traits (subjective attributes) and image aesthetic attributes (objective attributes) can be extracted simultaneously. In the second part, we leverage users' personality traits and image aesthetic attributes to jointly learn both aesthetic distribution and residual scores. In the third part, a target user's PIAA model is obtained by fine-tuning the prior model with a small number of annotated images. In the following subsections, we elaborate on these three parts.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

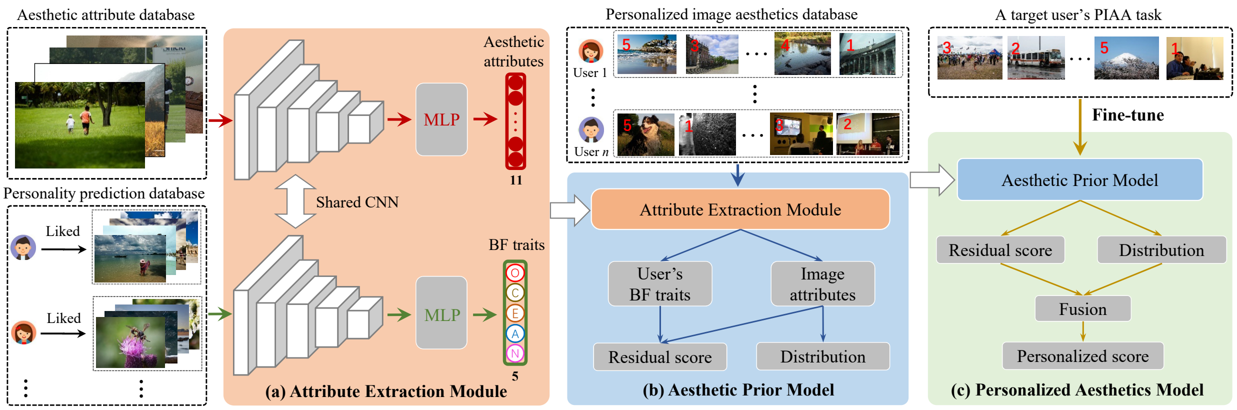


Fig. 2. An overview of the proposed PIAA-SOA framework, which contains three parts: an attribute extraction module, an aesthetic prior model, and a personalized aesthetics model. In the first part, users' personality traits and image aesthetic attributes can be extracted simultaneously. In the second part, we leverage users' personality traits and image aesthetic attributes to jointly learn both aesthetic distribution and residual scores. In the third part, a target user's PIAA model is obtained by fine-tuning the prior model with a small number of annotated images.

A. Attribute Extraction Module

Since we expect that our model can extract both the subjective attributes of users and the objective attributes of rated images simultaneously, an attribute extraction module is designed for the subsequent prior model. Inspired by multi-task learning [46], we utilize a shared convolutional neural network (CNN) and two multi-layer perceptrons (MLPs) to implement the proposed attribute extraction module, which is shown on the left side of Fig. 2. Specifically, the shared CNN is a popular deep network pre-trained on ImageNet [47], which can map an image x to the hidden features $\boldsymbol{d}$ and is formulated as

$$\boldsymbol{d} = f_\theta(x), \tag{1}$$

where θ denotes the parameters of the shared CNN f_θ. Then, we leverage two multi-layer perceptrons MLP_{θ_o} and MLP_{θ_s} to further map the hidden features $\boldsymbol{d}$ to the predicted objective attributes $\hat{\boldsymbol{o}}$ and subjective attributes $\hat{\boldsymbol{s}}$ respectively, which is defined as

$$\hat{\boldsymbol{o}} = MLP_{\theta_o}(\boldsymbol{d}), \hat{\boldsymbol{s}} = MLP_{\theta_s}(\boldsymbol{d}), \tag{2}$$

where θ_o and θ_s represent the parameters of MLP_{θ_o} and MLP_{θ_s} respectively, which are composed of two linear layers and a PReLU activation function. While objective attributes can be inferred from an image [15], the subjective attributes of users need to be mapped from a set of liked images [34]. Hence, the proposed attribute extraction module is jointly trained with the aesthetic attribute database [15] and the personality prediction database [34].

In this module, we assume that $\mathcal{D}_{aes} = \{x_i, \boldsymbol{o}_i\}_{i=1}^{N_a}$ and $\mathcal{D}_{per} = \{\mathcal{D}_{u_j}, \boldsymbol{s}_j\}_{i=j}^{N_p}$ denote the aesthetic attribute database and the personality prediction database, respectively, where $\boldsymbol{o}_i$ and $\boldsymbol{s}_j$ represent the labeled aesthetic attributes of image x_i ($i = 1, 2, 3, ..., N_a$) and the collected personality traits of user u_j ($j = 1, 2, 3, ..., N_p$). $\mathcal{D}_{u_j} = \{x_i^j\}_{i=1}^{N_{u_j}}$ indicates the subset with N_{u_j} images liked by the user u_j. Based on these two databases, we use the l_2 loss function to optimize the parameters θ, θ_o and θ_s of the attribute extraction module, which is defined as

$$\mathcal{L}_a = \frac{1}{N_a}\sum_{i=1}^{N_a}(\boldsymbol{o}_i - \hat{\boldsymbol{o}}_i)^2 + \frac{1}{N_p}\sum_{j=1}^{N_p}(\boldsymbol{s}_j - \hat{\boldsymbol{s}}_j)^2, \tag{3}$$

where $\hat{\boldsymbol{o}}_i$ and $\hat{\boldsymbol{s}}_j$ represent the predicted aesthetic attributes of image x_i and the predicted personality traits of user u_j, which can be computed by

$$\hat{\boldsymbol{o}}_i = MLP_{\theta_o}(f_\theta(x_i)), \tag{4}$$

$$\hat{\boldsymbol{s}}_j = \frac{1}{N_{u_j}}\sum_{i=1}^{N_{u_j}} MLP_{\theta_s}(f_\theta(x_i^j)). \tag{5}$$

In this manner, the attribute extraction module is obtained through joint training of $\mathcal{D}_{aes}$ and $\mathcal{D}_{per}$ and can simultaneously predict a user's BF personality traits from a set of liked images and extract the predicted aesthetic attributes of these images.

B. Aesthetic Prior Model

After attribute extraction, we need to capture users' personality traits and the aesthetic attributes of images, which can provide effective prior knowledge for the learning personalized image aesthetics of users. Consequently, we can refine the prior model from extensive individual users' personalized image aesthetic evaluation of images. Compared with the previous methods [25]–[30], the proposed method can embed subjective and objective attributes from the perspective of both users and images into the prior model, which explicitly characterizes the inherent factors that cause users' aesthetic preferences for images. In view of this, the PIAA of each user in the personalized image aesthetics database is regarded as an independent task and we obtain a large number of PIAA tasks for learning the aesthetic prior model, which is shown in the middle of Fig. 2.

In this model, we denote $\mathcal{D}_{prior} = \{\mathcal{D}_i\}_{i=1}^{n}$ as the personalized image aesthetics database, where $\mathcal{D}_i = \{x_{i,j}, y_{i,j}, \boldsymbol{d}_{i,j}\}_{j=1}^{m}$ represents the subset of the i-th user's PIAA task, where $y_{i,j}$ and $\boldsymbol{d}_{i,j}$ indicate the i-th user's per-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

sonalized score and the aesthetic distribution of the j-th rated image, respectively. Specifically, we fed all images of the subset $\mathcal{D}_i$ into the trained attribute extraction module, which takes the form

$$\hat{\boldsymbol{o}}_{i,j} = MLP_{\theta_o}(f_\theta(x_{i,j})), \hat{\boldsymbol{s}}_{i,j} = MLP_{\theta_s}(f_\theta(x_{i,j})), \tag{6}$$

where $\hat{\boldsymbol{o}}_{i,j}$ and $\hat{\boldsymbol{s}}_{i,j}$ denote the predicted aesthetic attributes of the j-th image rated by the i-th user and the predicted personality traits of users who like the image. To obtain the personality traits of the i-th user, we leverage the user's personalized score to weight the personality traits inferred from all m images, which is defined as

$$\hat{\boldsymbol{s}}_i = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{2(y_{i,j} - y_i^{med})}{y_i^{max} - y_i^{min}}\hat{\boldsymbol{s}}_{i,j}\right), \tag{7}$$

where y_i^{max}, y_i^{med}, and y_i^{min} represent the maximum, median and minimum values of the i-th user's personalized scores. In this way, we leverage users' personalized scores on images to transform image-level personality traits predicted by the attribute extraction module into user-level personality traits.

As mentioned above, aesthetic attributes determine the generic aesthetics of images (**distribution**), and the difference between users' personalized aesthetics and the generic aesthetics of images (**residual score**) is affected by both their personality traits and image aesthetic attributes. Since the input and output of the proposed aesthetic prior model are both one-dimensional feature vectors, two fully connected (FC) layers FC_{θ_d} and FC_{θ_r} are used to predict aesthetic distribution $\hat{\boldsymbol{d}}_{i,j}$ and residual score $\hat{r}_{i,j}$ respectively, which is formulated as

$$\hat{\boldsymbol{d}}_{i,j} = FC_{\theta_d}(\hat{\boldsymbol{o}}_{i,j}), \hat{r}_{i,j} = FC_{\theta_r}(\hat{\boldsymbol{s}}_i \otimes \hat{\boldsymbol{o}}_{i,j}), \tag{8}$$

where $\otimes$ is the Kronecker product, and θ_d and θ_r represent the parameters of FC_{θ_d} and FC_{θ_r} respectively. Considering that the sum of aesthetic distribution is 1, FC_{θ_d} contains two linear layers and a Softmax activation function. In contrast, since the residual score is a numerical value, FC_{θ_r} is composed of two linear layers and a PReLU activation function. Then, we adopt the l_2 loss function to optimize the parameters θ_r and θ_d of the proposed prior model, which is defined as

$$\mathcal{L}_p = \frac{1}{n}\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}((\boldsymbol{d}_{i,j} - \hat{\boldsymbol{d}}_{i,j})^2 + (r_{i,j} - \hat{r}_{i,j})^2), \tag{9}$$

where $r_{i,j}$ represents the residual score between the i-th user's personalized score and the generic score of the j-th image, which can be computed by

$$r_{i,j} = y_{i,j} - \text{mean}(\boldsymbol{d}_{i,j}), \tag{10}$$

where mean$(\cdot)$ indicates the average operation on the aesthetic distribution $\boldsymbol{d}_{i,j}$.

In this way, the proposed prior model can be trained by iteratively sampling the PIAA task of each user in $\mathcal{D}_{prior}$, which can firmly learn the prior knowledge that affects users' aesthetic preferences on images from both the personality traits of users and the aesthetic attributes of images simultaneously. The prior model can not only output the aesthetic distribution of images but also leverage users' subjective attributes to predict the residual scores of their personalized aesthetics relative to the generic aesthetics of images.

Algorithm 1 The proposed PIAA-SOA model

Input: Aesthetic attribute database $\mathcal{D}_{aes} = \{x_i, \boldsymbol{o}_i\}_{i=1}^{N_a}$, personality prediction database $\mathcal{D}_{per} = \{\mathcal{D}_{u_j}, \boldsymbol{s}_j\}_{i=j}^{N_p}$, personalized image aesthetics database $\mathcal{D}_{prior} = \{\mathcal{D}_i\}_{i=1}^{n}$, training set of a target user's PIAA task $\mathcal{D}_t = \{x_i, y_i\}_{i=1}^{m_t}$

Output: A PIAA model of the target user

Initialize all the parameters of the proposed model;
/⋆ Attribute Extraction Module ⋆/
for *iteration* $= 1, 2, ...$ **do**
  Sample a batch of k images from $\mathcal{D}_{aes}$, and N_{u_j} liked images of the user u_j from $\mathcal{D}_{per}$;
  for $j = 1, 2, ..., N_p$ **do**
    Output image aesthetic attributes $\{\hat{\boldsymbol{o}}_i\}_{i=1}^{k}$ and user's personality traits $\hat{\boldsymbol{s}}_j$ by using f_θ, MLP_{θ_o} and MLP_{θ_s};
    Update θ, θ_o and θ_s with the loss function $\mathcal{L}_a$;
  end for
end for
/⋆ Aesthetic Prior Model ⋆/
for *iteration* $= 1, 2, ...$ **do**
  Sample a user's subset $\mathcal{D}_i$ from $\mathcal{D}_{prior}$;
  for $i = 1, 2, ..., n$ **do**
    Compute the i-th user's personality traits $\hat{\boldsymbol{s}}_i$ and the aesthetic attributes $\hat{\boldsymbol{o}}_{i,j}$ of the j-th image;
    Output image aesthetic distribution $\boldsymbol{d}_{i,j}$ and residual score $r_{i,j}$ by using FC_{θ_d} and FC_{θ_r};
    Update θ_d and θ_r with the loss function $\mathcal{L}_p$;
  end for
end for
/⋆ Personalized Aesthetics Model ⋆/
Compute a target user's personality traits $\hat{\boldsymbol{s}}$ and image aesthetic attributes $\hat{\boldsymbol{o}}_i$;
Compute aesthetic distribution $\hat{\boldsymbol{d}}_i$ and residual score $\hat{r}_i$;
Output personalized score $\hat{y}_i$ by using FC_{θ_t};
Update θ_t with the loss function $\mathcal{L}_t$;
Obtain the target user's PIAA model.

C. Personalized Aesthetics Model

Since only a small number of images annotated by a target user can be obtained, we use these data to fine-tune the prior model to obtain the target user's PIAA model, which is shown on the right side of Fig. 2. We assume that $\mathcal{D}_t = \{x_i, y_i\}_{i=1}^{m_t}$ denotes the training set of a target user's PIAA task, where y_i represents the user's personalized aesthetic score of image x_i, and m_t is the number of annotated images. For the i-th image rated by the user, aesthetic attributes $\hat{\boldsymbol{o}}_i$ and personality traits $\hat{\boldsymbol{s}}_i$ can be predicted from the trained attribute extraction module, which takes the form

$$\hat{\boldsymbol{o}}_i = MLP_{\theta_o}(f_\theta(x_i)), \hat{\boldsymbol{s}}_i = MLP_{\theta_s}(f_\theta(x_i)). \tag{11}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

Particularly, we use $\{\hat{\boldsymbol{s}}_i\}_{i=1}^{m_t}$ of all m_t images to calculate the user's personality traits in the following form

$$\hat{\boldsymbol{s}} = \frac{1}{m_t}\sum_{i=1}^{m_t}\left(\frac{2(y_i - y^{med})}{y^{max} - y^{min}}\hat{\boldsymbol{s}}_i\right), \tag{12}$$

where y^{max}, y^{med}, and y^{min} represent the maximum, median and minimum values of the user's personalized scores. Then, the aesthetic distribution and residual score of the i-th image can be predicted from the trained prior model

$$\hat{\boldsymbol{d}}_i = FC_{\theta_d}(\hat{\boldsymbol{o}}_i), \hat{r}_i = FC_{\theta_r}(\hat{\boldsymbol{s}} \otimes \hat{\boldsymbol{o}}_i). \tag{13}$$

Finally, the user's predicted personalized score for the i-th image can be obtained by fusing the residual score $\hat{r}_i$ and aesthetic distribution $\hat{\boldsymbol{d}}_i$, which is defined as

$$\hat{y}_i = \hat{r}_i + FC_{\theta_t}(\hat{\boldsymbol{d}}_i), \tag{14}$$

where θ_t represents the parameters of the linear layer FC_{θ_t}, and it can be optimized by the following l_2 loss function

$$\mathcal{L}_t = \frac{1}{m_t}\sum_{i=1}^{m_t}(y_i - \hat{y}_i)^2. \tag{15}$$

In the fine-tuning process, FC_{θ_t} can automatically fuse the aesthetic distribution and residual scores into personalized scores through the learning of a few parameters θ_t. By fine-tuning the model parameters θ_t, we finally obtain a PIAA model that can simulate the user's personalized aesthetics for images. In summary, the whole optimization process of our algorithm is shown in Algorithm 1.

IV. Experiments

In our approach, we first use an aesthetic attribute database (AADB [15]) and a personality prediction database (PsychoFlickr [34]) to train the attribute extraction module. Then, the PIAA tasks of users in the training set of the FLICKR-AES database [25] are adopted to learn the aesthetic prior model. Finally, the PIAA model of a user is obtained by fine-tuning with the user's PIAA task in the testing set of the FLICKR-AES database and the REAL-CUR database [25].

A. Databases

The **AADB database** [15] collected 10,000 images evaluated by a total of 190 users. In addition to the aesthetic score, each image in this database also received 11 aesthetic attributes annotated by at least 5 users: *balancing element, color harmony, interesting content, depth of field, good lighting, motion blur, object emphasis, repetition, rule of thirds, symmetry*, and *vivid color*. The ranges of aesthetic scores and aesthetic attributes are [1, 5] and [-1, 1], respectively, and the higher the value is, the higher the aesthetics of images. Hence, we can use 11 aesthetic attributes of images in the AADB as the objective attribute labels to train the proposed attribute extraction module.

The **PsychoFlickr database** [34] collected 60,000 images liked by 300 users on the Flickr website, and each user had 200 liked images. Each user's BF personality traits, *Openness (O), Conscientiousness (C), Extroversion (E), Agreeableness (A)*, and *Neuroticism (N)*, were obtained by the BIF-10 questionnaire [43]. The BF personality traits range from -4 to 4, and a higher value indicates stronger personality traits. Therefore, users' BF personality traits can be used as the subjective attribute labels of their liked images to train the proposed attribute extraction module.

The **FLICKR-AES database** [25] collected 40,000 images evaluated by a total of 210 users. Among them, 35,263 images rated by 173 users were used as the training set, and the rest of the 4,737 images rated by 37 users were used as the testing set. The range of aesthetic scores is [1, 5], and the higher the value, the higher the aesthetics of images. Hence, we can leverage 173 users' PIAA tasks in the training set to train the aesthetic prior model. In the testing set, the number of images rated by each user is in the range of 110 and 190. Then, the PIAA task of a user in the testing set is used to fine-tune the prior model to obtain the PIAA model that can automatically predict this user's personalized aesthetic score for images.

The **REAL-CUR database** [25] collected personal photo albums of 14 users and their aesthetic ratings on images in their albums. The number of images in these users' photo albums is in the range of 197 to 222. The range of aesthetic ratings is [1, 5], and a higher value corresponds to higher image aesthetics. Therefore, the PIAA model of each user can be obtained by fine-tuning the prior model trained on the training set of the FLICKR-AES database. This database can verify the performance of a user's PIAA model when only a few annotated images are available in real scenarios.

B. Experimental Settings

1) Implementation Details: In the proposed framework, ResNet18 [48] and Inception-v3 [49] are used as the shared CNN that is pre-trained on ImageNet [47]. MLP_{θ_o} consists of two linear layers with 512 nodes and 11 nodes, and MLP_{θ_s} consists of two linear layers with 512 nodes and 5 nodes. The two linear layers of FC_{θ_d} contain 64 nodes and 5 nodes, and the two linear layers of FC_{θ_r} contain 64 nodes and 1 node. The linear layer FC_{θ_t} is composed of 5 nodes. All the parameters of these linear layers are randomly initialized. In the attribute extraction module, we set the learning rates of the shared CNN and linear layers to $1e{-}5$ and $1e{-}3$, respectively. The batch of k images and N_{u_j} liked images are set to 100 and 200, respectively. In the aesthetic prior model, the learning rate of FC_{θ_d} and FC_{θ_r} is set to $1e{-}3$. In the PIAA model for a specific user, we set the learning rate of FC_{θ_t} to $1e-5$. In our experiment, the maximum, median, and minimum values of users' personalized scores are 5, 3, and 1, respectively. In the training of the attribute extraction module and the aesthetic prior model, the total number of epochs is 100. In the fine-tuning of a PIAA model, we set the number of epochs to 50. All the learning rates drop to a factor of 0.5 after every 10 epochs. We adopt Adam [50] to optimize the parameters of our model, which is implemented based on PyTorch [51] and a machine with a NVIDIA GeForce GTX 1080 Ti.

2) Compared Methods: At present, several PIAA methods [25], [26], [28]–[30], [40] have released their testing results on public PIAA databases. We compare the proposed PIAA-SOA with these six representative PIAA methods, which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

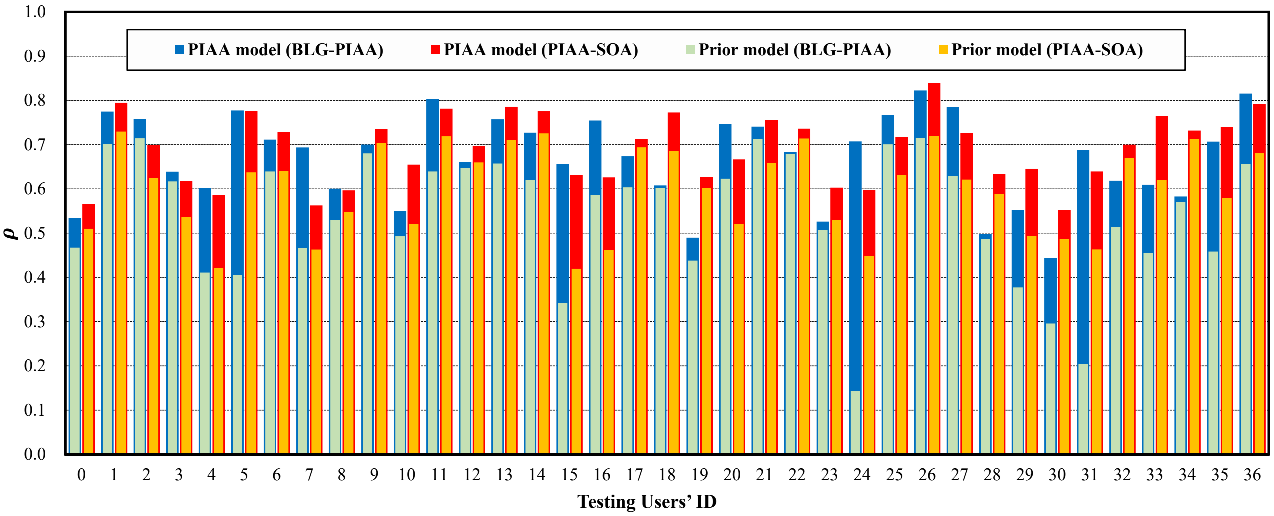


Fig. 3. Evaluation results (ρ) of 37 testing users on the FLICKR-AES database [25] by using the prior model and PIAA model of BLG-PIAA [30] and PIAA-SOA. The green and blue bars show the testing results of BLG-PIAA, and the orange and red bars show the testing results of PIAA-SOA.

TABLE I
COMPARISON RESULTS (AVERAGE ρ) OF THE PROPOSED PIAA-SOA AND THE STATE-OF-THE-ART METHODS ON THE FLICKR-AES DATABASE.

Method	10 images	100 images
FPMF (attribute) [40]	0.511±0.004	0.516±0.003
FPMF (content) [40]	0.512±0.002	0.516±0.010
FPMF (content and attribute) [40]	0.513±0.003	0.524±0.007
USAR_PPR [26]	0.521±0.002	0.544±0.007
USAR_PAD [26]	0.520±0.003	0.537±0.003
USAR_PPR&PAD [26]	0.525±0.004	0.552±0.015
PAM (attribute) [25]	0.518±0.003	0.539±0.013
PAM (content) [25]	0.515±0.004	0.535±0.017
PAM (content and attribute) [25]	0.520±0.003	0.553±0.012
Wang *et al.* [28]	0.522±0.005	0.562±0.015
PA_IAA [29]	0.543±0.003	0.639±0.011
BLG-PIAA [30]	0.561±0.005	0.669±0.013
PIAA-SOA	**0.618±0.006**	**0.691±0.015**

include a collaborative filtering-based approach (FPMF [40]), a user interaction-based approach (USAR [26]) and four aesthetic difference-based approaches (PAM [25], Wang *et al.* [28], PA_IAA [29] and BLG-PIAA [30]).

3) Evaluation Criterion: Since people's aesthetic evaluation of images is relatively subjective, the ranking correlation between predicted aesthetic scores and ground-truth aesthetic scores is an effective evaluation criterion [25], [26]. Therefore, the Spearman rank-order correlation coefficient (ρ) is employed to evaluate the performance of PIAA models. The values of ρ range from -1 to 1, and better PIAA methods should have larger values of ρ.

C. Performance Evaluation on PIAA Models

To evaluate the performance of our PIAA-SOA model for learning the personalized image aesthetics of individual users, we compare PIAA-SOA with six state-of-the-art PIAA methods on 37 testing users of the FLICKR-AES database [25]. For each user, m_t annotated images are used to fine-tune the aesthetic prior model to obtain a PIAA model. For a fair comparison with the reported results of existing methods [25], [26], [28]–[30], m_t is set to 10 and 100 in this experiment. To avoid random bias, we perform 50 repeated experiments and report the average results and the corresponding standard deviation for each user.

In Table I, we average the results (ρ) of 37 testing users in the FLICKR-AES database and list the comparison results (average ρ) of the proposed PIAA-SOA and the state-of-the-art methods, where the best results are shown in bold. In this experiment, we adopt ResNet18 [48] as the shared CNN of our model. As shown in the table, our PIAA-SOA method significantly outperforms the collaborative filtering-based method (FPMF [40]) and the user interaction-based method (USAR [26]), possibly because our method can explicitly infer users' subjective attributes by merely using their aesthetic ratings on images without redundant user interactions and aesthetic correlation calculations among different users. Compared with four aesthetic difference-based methods (PAM [25], Wang *et al.* [28], PA_IAA [29] and BLG-PIAA [30]), our method also yields superior performance in learning individual users' personalized image aesthetics, indicating that our method can embed the personality traits of users and the aesthetic attributes of images into the proposed aesthetic prior model by learning the PIAA tasks of extensive users, which is more efficient than learning the aesthetic prior model only from the objective attributes of images.

To further verify whether the proposed PIAA-SOA model can effectively learn the personalized image aesthetics of each user, we compare the proposed PIAA-SOA with the modern advanced BLG-PIAA [30] and evaluate the prior model and PIAA model ($m_t = 100$) of these two methods. In our prior model, the average result of the image aesthetic distribution is used as the predicted aesthetic score. Fig. 3 shows the evaluation results of 37 testing users in terms of ρ. From the figure, we observe that the proposed PIAA-SOA is superior

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

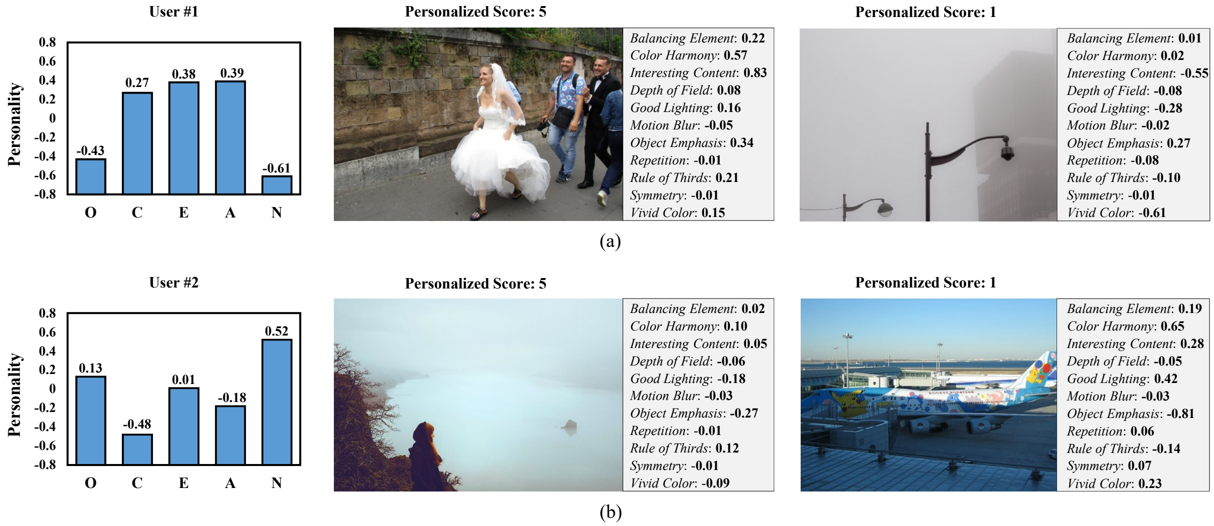


Fig. 4. Two users' BF personality traits and image aesthetic attributes predicted by the proposed attribute extraction module. The two users are from the training set of the FLICKR-AES database [25]. Besides, two images rated by each user and the corresponding personalized scores are also shown. (a) User #1. (b) User #2.

to BLG-PIAA in both the prior model and the PIAA model for most users (21 out of 37). Specifically, our prior model can achieve better performance than BLG-PIAA for each user (0.598 versus 0.540), which demonstrates the effectiveness of fusing the personality traits of users and the aesthetic attributes of images in the proposed prior model. In addition, when the proposed PIAA-SOA model is fine-tuned on users' small amount of personalized data, the increase of average ρ for 37 testing users is about 0.093 (from 0.598 to 0.691). This proves that our prior model can be easily adapted to an effective PIAA model for individual users.

D. Effectiveness of Attribute Extraction Module

To verify the effectiveness of the proposed attribute extraction module in capturing users' personality traits and image aesthetic attributes, Fig. 4 shows the qualitative results of two example users in the training set of the FLICKR-AES database [25]. The predicted BF personality traits of these two users can be obtained by Eq. 7, which are normalized into the range $[-1, 1]$. In addition, two images rated by each user and the corresponding predicted aesthetic attributes are also shown. As shown in the figure, User #1 is a person with high extroversion and agreeableness. Hence, he has higher aesthetic scores for images with interesting content and harmonious color. In contrast, he tends to assign lower aesthetic scores to images with monotonous content and color. User #2 is a neurotic person and prefers images with dark and quiet scenes. As a result, he has lower aesthetic scores for bright and colorful images. The results on the attribute extraction module have shown good performance in characterizing the subjective attributes of individual users and the objective attributes of their rated images, which will provide solid knowledge to learn the aesthetic prior model.

Based on the above analysis, users' personality traits may have a strong correlation with their preferred image aesthetic attributes. To further explore the relationship between users'

	O	*C*	*E*	*A*	*N*
Balancing Element	0.7339	0.8773	0.7143	0.8854	-0.7716
Color Harmony	0.8396	0.9608	0.8355	0.9638	-0.7596
Interesting Content	0.7790	0.9223	0.7753	0.9296	-0.7729
Depth of Field	0.5195	0.6332	0.5549	0.6547	-0.5756
Good Lighting	0.5794	0.7450	0.6886	0.7839	-0.7248
Motion Blur	-0.7503	-0.7287	-0.7845	-0.7068	0.3651
Object Emphasis	0.1599	0.1870	0.2014	0.2156	-0.2672
Repetition	0.8641	0.8624	0.7726	0.8382	-0.5969
Rule of Thirds	0.6963	0.5680	0.7421	0.5324	-0.1683
Symmetry	0.7912	0.8436	0.6361	0.8177	-0.6625
Vivid Color	0.5323	0.6241	0.7674	0.6714	-0.5255

Fig. 5. The ranking correlation (ρ) in each dimension of BF personality traits and 11 aesthetic attributes on the training set of the FLICKR-AES database [25]. The red background shows a positive ranking correlation, and the green background shows a negative ranking correlation.

subjective attributes and image objective attributes, we conduct a quantitative correlation analysis from the perspective of users. Similar to Eq. 7, we utilize all the images rated by a user to calculate his preferred aesthetic attributes. Particularly, the i-th user preferred image aesthetic attributes $\hat{\boldsymbol{o}}_i$ are formulated as

$$\hat{\boldsymbol{o}}_i = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{2(y_{i,j} - y_i^{med})}{y_i^{max} - y_i^{min}} \hat{\boldsymbol{o}}_{i,j} \right), \tag{16}$$

where m denotes the number of images in the subset of a user's PIAA task, and y_i^{max}, y_i^{med}, and y_i^{min} represent the maximum, median and minimum values of the i-th user's personalized scores. Based on the trained attribute extraction module, we leverage ρ to investigate the ranking correlation in each dimension of BF personality traits and 11 aesthetic attributes.

TABLE II
COMPARISON RESULTS (AVERAGE ρ) OF OUR PIAA-SOA, PA_IAA AND BLG-PIAA ON TESTING USERS IN THE AADB AND REAL-CUR DATABASES.

Database	Method	10 images	100 images
AADB	PA_IAA [29]	0.469±0.002	0.524±0.006
	BLG-PIAA [30]	0.486±0.004	0.536±0.006
	PIAA-SOA	**0.509±0.003**	**0.557±0.007**
REAL-CUR	PA_IAA [29]	0.443±0.004	0.562±0.013
	BLG-PIAA [30]	0.448±0.007	0.578±0.015
	PIAA-SOA	**0.487±0.006**	**0.589±0.014**

Fig. 5 shows the experimental results for 173 training users in the FLICKR-AES database [25]. From the figure, most of the aesthetic attributes preferred by users have a strong correlation with their BF personality traits. The aesthetic attributes preferred by users with high neuroticism are opposite to those of users with other high personality traits. Specifically, neuroticism is negatively correlated with most aesthetic attributes, and the other four personality traits are positively correlated with most aesthetic attributes, which confirms the qualitative experimental results in Fig. 4. Therefore, users' aesthetic preferences for images are determined by both their personality traits and image aesthetic attributes, and the stable correlation between the subjective and objective attributes is explicit prior knowledge that can be learned in our aesthetic prior model.

E. Cross Database Evaluation

To validate the generalization performance of our prior model learned from the FLICKR-AES database [25], we conduct a cross-database evaluation on the testing users of AADB [15] and REAL-CUR [25]. Although AADB is an aesthetic attribute database, it can be used to evaluate the performance of PIAA methods because it also provides users' identity information. Similar to the FLICKR-AES database, we select users whose number of annotated images ranges from 110 to 200 for model testing and obtain a total of 22 users in the AADB database. For each testing user in AADB and REAL-CUR, we randomly use m_t annotated images to fine-tune the prior model trained on the FLICKR-AES database. In this experiment, we set m_t to 10 and 100. To avoid bias, the average results and the corresponding standard deviation of 50 repeated experiments are reported. Since only the source codes of PA_IAA [29] and BLG-PIAA [30] are available to us, we compare PIAA-SOA with these two methods under the above experimental settings. We average the results (ρ) of all testing users in each database and summarize the comparison results of these three methods in Table II. For each database, the best results are shown in bold. The figure shows that our method outperforms the other two methods on both databases. Specifically, the test results of our prior model (PIAA-SOA) on AADB and REAL-CUR are 0.476 and 0.461, respectively. When only 10 annotated images are adopted to fine-tune the prior model, the average ρ of our PIAA-SOA model for all testing users has a certain improvement (0.033 for AADB and 0.026 for REAL-CUR), indicating that the prior model of our PIAA-SOA can be used as more generalized prior knowledge for learning the personalized image aesthetics of users. In particular, the experimental results on REAL-CUR show that our method can provide reliable prior knowledge for learning a user's PIAA model in practical applications.

TABLE III
ABLATION STUDY RESULTS (AVERAGE ρ) OF THE PROPOSED MODEL ON THE FLICKR-AES DATABASE.

Backbone	Model	10 images	100 images
ResNet18	PIAA-SOA w/o attr	0.542±0.003	0.618±0.011
	PIAA-SOA w/o trait	0.534±0.004	0.612±0.009
	PIAA-SOA w/o residual	0.566±0.003	0.639±0.012
	PIAA-SOA w/o fusion	0.609±0.005	0.680±0.014
	PIAA-SOA	**0.618±0.006**	**0.691±0.015**
Inception-v3	PIAA-SOA w/o attr	0.538±0.005	0.621±0.009
	PIAA-SOA w/o trait	0.522±0.003	0.609±0.010
	PIAA-SOA w/o residual	0.569±0.004	0.634±0.011
	PIAA-SOA w/o fusion	0.602±0.006	0.676±0.013
	PIAA-SOA	**0.611±0.005**	**0.688±0.017**

F. Ablation Study

In this section, we perform an ablation study to verify the contribution of each module in our model for learning the personalized image aesthetics of individual users. In the attribute extraction module, we remove the aesthetic attributes branch (PIAA-SOA w/o attr) and the personality traits branch (PIAA-SOA w/o trait). In the aesthetic prior model, we replace the distribution and residual scores with personalized scores, which is called "PIAA-SOA w/o residual". In the personalized aesthetics model, we remove the fusion module and directly fine-tune the prior model (PIAA-SOA w/o fusion). In addition, we adopt two backbone networks (ResNet18 [48] and Inception-v3 [49]) to conduct these ablation experiments on the FLICKR-AES database [25] and summarize the tested results in Table III, where the best result on each backbone is highlighted in bold font.

As shown in Table III, the full version of our PIAA-SOA model yields the best evaluation performance on both backbone networks. PIAA-SOA outperforms "PIAA-SOA w/o attr" and "PIAA-SOA w/o trait" by a large margin, which indicates that aesthetic attributes and personality traits can effectively characterize the objective attribute of images and the subjective attribute of users in our PIAA model. When the learning of distribution and residual scores is replaced by only learning personalized scores (PIAA-SOA w/o residual), the performance of our model is also degraded, demonstrating that the aesthetic differences among different users can be effectively modeled by the personality traits of users and the aesthetic attributes of images. Compared with "PIAA-SOA w/o fusion", PIAA-SOA also achieves slight performance improvement, which illustrates the effectiveness of automatically fusing the distribution and residual score when fine-tuning the prior model with a small amount of annotated data from users. In summary, each of the above modules contributes to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

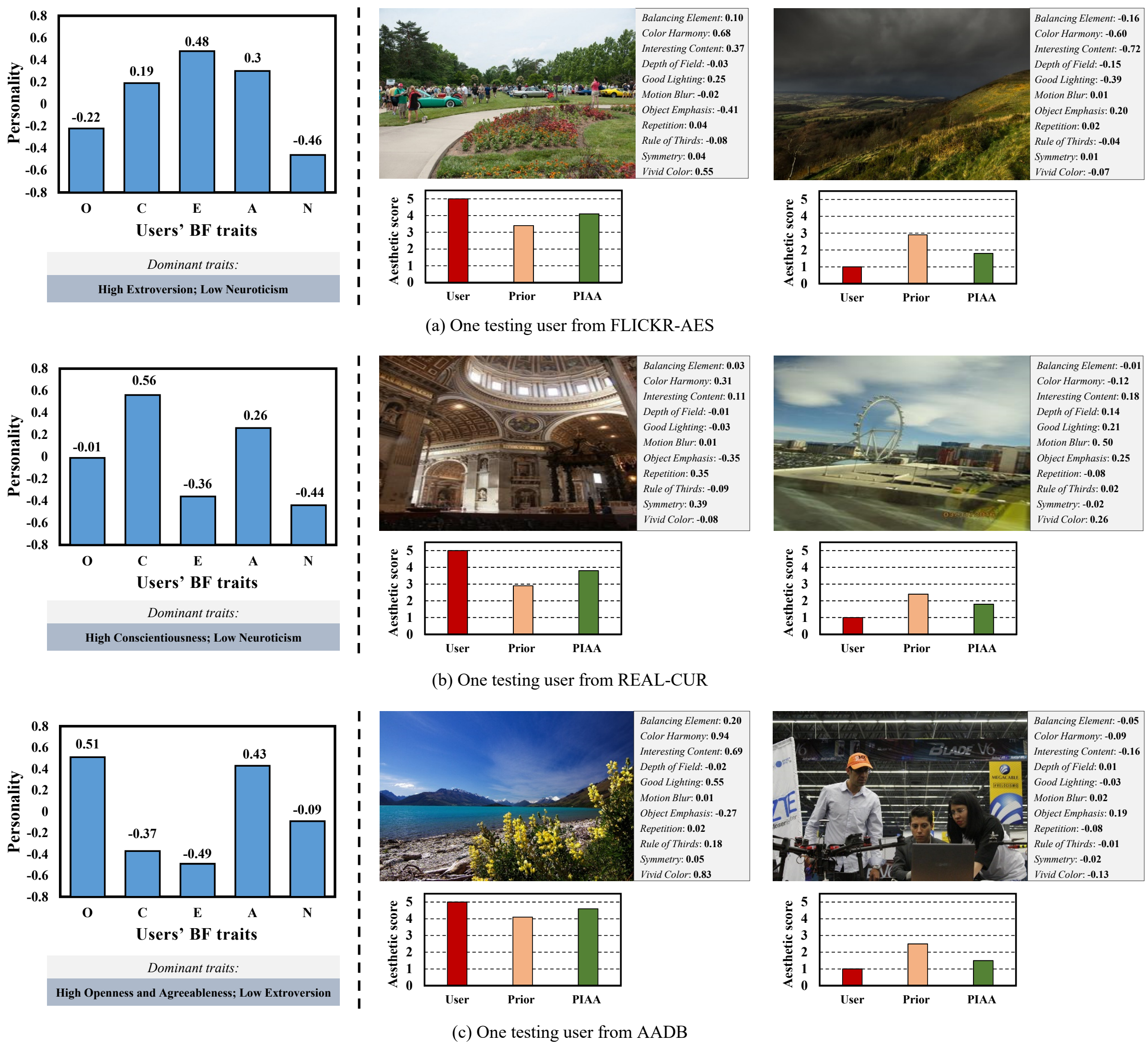


Fig. 6. Example results of our PIAA-SOA model on three testing users from the FLICKR-AES, AADB, and REAL-CUR databases. The left side shows the predicted BF personality traits of each user, which are normalized into the range [-1, 1]. The predicted 11 aesthetic attributes are shown on the right side of each image. The personalized aesthetic scores rated by users and the predicted aesthetic scores of our prior model and PIAA model are shown below each image.

the proposed model, and ResNet18 is used as the backbone network in our experiments.

G. Visual Analysis

To further show the effectiveness of our method and how the proposed PIAA-SOA model works, we introduce a visual experiment on three testing users from the FLICKR-AES, AADB, and REAL-CUR databases. In the experiment, 100 images ($m_t = 100$) annotated by each testing user are used to obtain the PIAA model. Example results of our PIAA-SOA model on the three users are shown in Fig. 6. As shown in the figure, the attribute extraction module can effectively predict users' personality traits and the aesthetic attributes of images. For the three users, the aesthetic score predicted by the PIAA model is more accurate than that predicted by the prior model, indicating that the proposed prior model is fine-tuned with a small amount of annotated data to obtain a more precise PIAA model for a user. In particular, the dominant traits of the user in Fig. 6(a) are high extroversion and low neuroticism, and he tends to assign a higher aesthetic score to the image with bright and colorful outdoor scenes. In contrast, he thinks that the image with bad light is not beautiful. As shown in Fig. 6(b), the user is a person with high conscientiousness and low neuroticism, which shows that he is orderly and cautious and prefers the image with symmetry and repetition. Fig. 6(c) shows that the user is an open, friendly, and introverted person, indicating that he assigns a higher aesthetic rating to the natural image with colorful, pleasant, and interesting content. Consequently, he does not like the image with serious and boring content. As shown by the above visual analysis, our PIAA model leverages the stable correlation between the subjective attributes of users and the objective attributes of images as prior knowledge to effectively learn the personalized image aesthetics of individual users.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3123468, IEEE Transactions on Multimedia

V. Conclusion

This paper has presented a personalized image aesthetics assessment model based on subjective and objective attributes (PIAA-SOA). Different from existing PIAA models, the proposed PIAA-SOA model can characterize the subjective and objective attributes that determine image aesthetics from the perspective of both users and images simultaneously. The proposed attribute extraction module has demonstrated its effectiveness in capturing the aesthetic attributes of images and the BF personality traits of users. Based on the PIAA tasks of extensive users, our method effectively embeds the stable correlation between users' personality traits and the aesthetic attributes of their rated images into the aesthetic prior model. Consequently, when a specific user can provide only a small number of annotated images, the proposed aesthetic prior model can be easily adapted to the PIAA model that conforms to the personalized image aesthetics of the user. Extensive experimental results and visual analysis have shown that the proposed PIAA-SOA model can not only effectively extract the subjective attributes of users and the objective attributes of images in visual aesthetics, but is also efficient in learning users' personalized image aesthetics.

Although our method has made significant progress in learning the personalized image aesthetics of users, the overall evaluation performance is still moderate (the ranking correlation coefficient on the FLICKR-AES database is 0.691). This is mainly because the subjective attributes that affect users' aesthetic preferences are complex, and more subjective attributes besides personality traits need to be considered (e.g. emotion, gender, age, and cultural background). For example, a user's transience emotion may change his original aesthetic experience [52]. Besides, the personality traits of users captured by our method strictly rely on a small amount of their annotated aesthetic data, which will make the obtained personality traits unstable. In view of this, it is necessary to consider more subjective attributes that affect users' visual aesthetics in the PIAA model. To obtain stable subjective attributes of users, there is an urgent need to develop more PIAA databases that incorporate diversified subjective attributes annotated by individual users in future work.

References

[1] Y. Deng, C. L. Chen, and X. Tang, "Image aesthetic assessment: An experimental survey," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 80–106, Jul. 2017.

[2] G. Guo, H. Wang, C. Shen, Y. Yan, and H. M. Liao, "Automatic image cropping for visual aesthetic enhancement using deep neural networks and cascaded regression," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2073–2085, Aug. 2018.

[3] P. Lu, H. Zhang, X. Peng, and X. Jin, "Learning the relation between interested objects and aesthetic region for image cropping," *IEEE Trans. Multimedia*, 2020, DOI:10.1109/TMM.2020.3029882.

[4] W. Sun, T. Chao, Y. Kuo, and W. H. Hsu, "Photo filter recommendation by category-aware aesthetic learning," *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1870–1880, Aug. 2017.

[5] J. Wang, S. Zhu, J. Xu, and D. Cao, "The retrieval of the beautiful: Self-supervised salient object detection for beauty product retrieval," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2548–2552.

[6] K. Karlsson, W. Jiang, and D.-Q. Zhang, "Mobile photo album management with multiscale timeline," in *Proc. ACM Int. Conf. Multimedia*, Nov. 2014, pp. 1061–1064.

[7] O. Wu, H. Zuo, W. Hu, and B. Li, "Multimodal web aesthetics assessment based on structural svm and multitask fusion learning," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1062–1076, Jun. 2016.

[8] S. E. Palmer, K. B. Schloss, and J. Sammartino, "Visual aesthetics and human preference," *Annu. Rev. Psychol*, vol. 64, pp. 77–107, 2013.

[9] M. Kucer, A. C. Loui, and D. W. Messinger, "Leveraging expert feature knowledge for predicting image aesthetics," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5100–5112, Oct. 2018.

[10] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: a large-scale database for aesthetic visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2408–2415.

[11] X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1930–1943, Dec. 2013.

[12] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rating image aesthetics using deep learning," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2021–2034, Nov. 2015.

[13] Y. Kao, R. He, and K. Huang, "Deep aesthetic quality assessment with semantic information," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1482–1495, Mar. 2017.

[14] H.-J. Lee, K.-S. Hong, H. Kang, and S. Lee, "Photo aesthetics analysis via dcnn feature encoding," *IEEE Trans. on Multimedia*, vol. 19, no. 8, pp. 1921–1932, Aug. 2017.

[15] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *Proc. Eur. Conf. Comput. Vis.*, May 2016, pp. 662–679.

[16] B. Pan, S. Wang, and Q. Jiang, "Image aesthetic assessment assisted by attributes through adversarial learning," in *Proc. AAAI Int. Conf. Artif. Intell.*, Jan. 2019, pp. 679–686.

[17] J.-T. Lee and C.-S. Kim, "Image aesthetic assessment based on pairwise comparison: A unified approach to score regression, binary classification, and personalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 1191–1200.

[18] W. H. Kim, J. H. Choi, and J. S. Lee, "Objectivity and subjectivity in aesthetic quality assessment of digital photographs," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 493–506, Jul. 2020.

[19] H. Talebi and P. Milanfar, "NIMA: neural image assessment," *IEEE Trans. on Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.

[20] C. Cui, H. Liu, T. Lian, L. Nie, L. Zhu, and Y. Yin, "Distribution-oriented aesthetics assessment with semantic-aware hybrid network," *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1209–1220, May 2019.

[21] X. Zhang, X. Gao, W. Lu, and L. He, "A gated peripheral-foveal convolutional neural network for unified image aesthetic prediction," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2815–2826, Nov. 2019.

[22] D. She, Y.-K. Lai, G. Yi, and K. Xu, "Hierarchical layout-aware graph convolutional network for unified aesthetics assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 8475–8484.

[23] X. Song, X. Han, Y. Li, J. Chen, X.-S. Xu, and L. Nie, "GP-BPR: personalized compatibility modeling for clothing matching," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2019, pp. 320–328.

[24] H. Kim, Y. J. Koh, and C. Kim, "PieNet: personalized image enhancement network," in *Proc. Eur. Conf. Comput. Vis.*, Aug. 2020, pp. 374–390.

[25] J. Ren, X. Shen, Z. Lin, R. Mech, and D. J. Foran, "Personalized image aesthetics," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jan. 2017, pp. 638–647.

[26] P. Lv, M. Wang, Y. Xu, Z. Peng, J. Sun, S. Su, B. Zhou, and M. Xu, "USAR: an interactive user-specific aesthetic ranking framework for images," in *Proc. ACM Int. Conf. Multimedia*, Nov. 2018, pp. 1328–1336.

[27] G. Wang, J. Yan, and Z. Qin, "Collaborative and attentive learning for personalized image aesthetic assessment," in *Proc. Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 957–963.

[28] W. Wang, J. Su, L. Li, X. Xu, and J. Luo, "Meta-learning perspective for personalized image aesthetics assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sept. 2019, pp. 1875–1879.

[29] L. Li, H. Zhu, S. Zhao, G. Ding, and W. Lin, "Personality-assisted multi-task learning for generic and personalized image aesthetics assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 3898–3910, 2020.

[30] H. Zhu, L. Li, J. Wu, S. Zhao, G. Ding, and G. Shi, "Personalized image aesthetics assessment via meta-learning with bilevel gradient optimization," *IEEE Trans. Cybern.*, 2020, DOI:10.1109/TCYB.2020.2984670.

[31] F. Gelli, T. Uricchio, X. He, A. D. Bimbo, and T. Chua, "Learning subjective attributes of images from auxiliary sources," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2263–2271.

[32] G. Matthews and I. J. Deary, "Personality traits," *Cambr. Univ. Press*, 2009.

[33] V. Swami and A. Furnham, "Personality and aesthetics preferences." Cambridge Univ. Press, 2014.

[34] C. Segalin, A. Perina, M. Cristani, and A. Vinciarelli, "The pictures we like are our image: Continuous mapping of favorite pictures into self-assessed and attributed personality traits," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 268–285, Apr. 2017.

[35] H. Zhu, L. Li, S. Zhao, and H. Jiang, "Evaluating attributed personality traits from scene perception probability," *Pattern Recognit. Lett.*, vol. 116, pp. 121–126, Dec. 2018.

[36] C. Cui, P. Lin, X. Nie, M. Jian, and Y. Yin, "Social-sensed image aesthetics assessment," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 3s, Dec. 2021.

[37] C. Cui, W. Yang, C. Shi, M. Wang, X. Nie, and Y. Yin, "Personalized image quality assessment with social-sensed aesthetic preference," *Inf. Sci.*, vol. 512, pp. 780–794, 2020.

[38] X. Deng, C. Cui, H. Fang, X. Nie, and Y. Yin, "Personalized image aesthetics assessment," in *Proc. ACM Conf. Inf. and Knowl. Manag.* ACM, Nov. 2017, pp. 2043–2046.

[39] K. Park, S. Hong, M. Baek, and B. Han, "Personalized image aesthetic quality assessment by joint regression and ranking," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, May 2017, pp. 1206–1214.

[40] P. O'Donovan, A. Agarwala, and A. Hertzmann, "Collaborative filtering of color aesthetics," in *Proc. Workshop Comput. Aesthet.*, Aug. 2014, pp. 33–40.

[41] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 273–291, Nov. 2014.

[42] R. R. McCrae, "The five-factor model of personality traits: Consensus and controversy," *Cambr. Handb. Psychol. Psychol.*, pp. 148–161, 2009.

[43] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german," *J. Res. Pers.*, vol. 41, no. 1, pp. 203–212, Feb. 2007.

[44] S. C. Guntuku, J. T. Zhou, S. Roy, W. Lin, and I. W. Tsang, "'who likes what and, why?' insights into modeling users' personality based on image 'likes'," *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 130–143, Jan. 2018.

[45] H. Zhu, L. Li, H. Jiang, and A. Tan, "Inferring personality traits from attentive regions of user liked images via weakly supervised dual convolutional network," *Neural Process. Lett.*, vol. 51, no. 3, pp. 2105–2121, Jun 2020.

[46] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, Jul. 1997.

[47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2012, pp. 1097–1105.

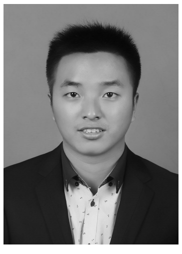
[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Nov. 2015, pp. 770–778.

[49] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, May 2015, pp. 1–15.

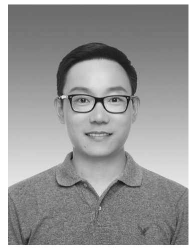
[51] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Proc. Adv. Neural Inf. Process. Syst. Workshop*, Dec. 2017, pp. 1–4.

[52] S. V. de Cruys and J. Wagemans, "Putting reward in art: A tentative prediction error account of visual art," *i-Perception*, vol. 2, no. 9, pp. 1035–1062, 2011.



Hancheng Zhu received the B.S. degree from Changzhou Institute of Technology, Changzhou, China, in 2012, and the M.S. and Ph.D. degrees from China University of Mining and Technology, Xuzhou, China, in 2015 and 2020, respectively. He is currently a Postdoctoral Fellow in the School of Computer Science and Technology, China University of Mining and Technology, China. His research interests include image aesthetics assessment and affective computing.

Yong Zhou received the B.S. degree from Hohai University, Nanjing, China, in 1997, and the M.S. and Ph.D. degree from China University of Mining and Technology, Xuzhou, China, in 2003 and 2006, respectively. From 2003 to present, he worked in the School of Computer Science and Technology, China University of Mining and Technology, as Assistant Professor, Associate Professor and Professor, respectively. His research mainly focuses on data mining, machine learning, computer vision and artificial intelligence.



Leida Li (Member, IEEE) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. In 2008, he was a Research Assistant with the Department of Electronic Engineering, Kaohsiung University of Science and Technology, Kaohsiung, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search (ROSE) Lab, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he was a Senior Research Fellow from 2016 to 2017. From 2009 to 2019, he worked in the School of Information and Control Engineering, China University of Mining and Technology, as Assistant Professor, Associate Professor and Professor. Currently, he is a Professor with the School of Artificial Intelligence, Xidian University.

His research interests include multimedia quality assessment, affective computing, information hiding, and image forensics. He has served as an SPC for IJCAI 2019-2021, the Session Chair for ICMR 2019 and PCM 2015, and a TPC for CVPR 2021, ICCV 2021, AAAI 2019-2021, ACM MM 2019-2020, ACM MM-Asia 2019, and ACII 2019. He is an Associate Editor of the Journal of Visual Communication and Image Representation and the EURASIP Journal on Image and Video Processing.

Yaqian Li received the B.S. degree from Lanzhou University, and the M.S. degree from Harbin Institute of Technology, China, in 2011 and 2013, receptively. He is currently the Technical Lead of visual search and understanding with OPPO Research Institute. His professional interests lie in the broad area of visual recognition, image retrieval, object detection, image quality assessment, and multi-modality learning.

Yandong Guo received the B.S. and M.S. degrees in ECE from Beijing University of Posts and Telecommunications, China, in 2005 and 2008, receptively, and the Ph.D. degree in ECE from Purdue University at West Lafayette in 2013, under the supervision of Prof. Bouman and Prof. Allebach.

He is currently the Chief Scientist of Intelligent Perception with OPPO and chair the AI strategic planning for OPPO. He also holds an adjunct professor position at the Beijing University of Posts and Telecommunications. Before he joined OPPO in 2020, he was the Chief Scientist with XPeng Motors, China, and previously a researcher with Microsoft Research, Redmond, WA, USA. His professional interests lie in the broad area of computer vision, imaging systems, human behavior understanding and biometric, and autonomous driving.