# Personalized Image Aesthetics Assessment via Meta-Learning With Bilevel Gradient Optimization

Hancheng Zhu, Leida Li, Jinjian Wu, Sicheng Zhao, Guiguang Ding, and Guangming Shi

***Abstract*—Typical image aesthetics assessment (IAA) is modeled for the generic aesthetics perceived by an "average" user. However, such generic aesthetics models neglect the fact that users' aesthetic preferences vary significantly depending on their unique preferences. Therefore, it is essential to tackle the issue for personalized IAA (PIAA). Since PIAA is a typical small sample learning (SSL) problem, existing PIAA models are usually built by fine-tuning the well-established generic IAA (GIAA) models, which are regarded as prior knowledge. Nevertheless, this kind of prior knowledge based on "average aesthetics" fails to incarnate the aesthetic diversity of different people. In order to learn the shared prior knowledge when different people judge aesthetics, that is, learn how people judge image aesthetics, we propose a PIAA method based on meta-learning with bilevel gradient optimization (BLG-PIAA), which is trained using individual aesthetic data directly and generalizes to unknown users quickly. The proposed approach consists of two phases: 1) meta-training and 2) meta-testing. In meta-training, the aesthetics assessment of each user is regarded as a task, and the training set of each task is divided into two sets: 1) support set and 2) query set. Unlike traditional methods that train a GIAA model based on average aesthetics, we train an aesthetic meta-learner model by bilevel gradient updating from the support set to the query set using many users' PIAA tasks. In meta-testing, the aesthetic meta-learner model is fine-tuned using a small amount of aesthetic data of a target user to obtain the PIAA model. The experimental results show that the proposed method outperforms the state-of-the-art PIAA metrics, and the learned prior model of BLG-PIAA can be quickly adapted to unseen PIAA tasks.**

***Index Terms*—Bilevel gradient optimization, meta-learning, personalized image aesthetics assessment (PIAA), small sample learning (SSL).**

Manuscript received September 9, 2019; revised February 9, 2020; accepted March 27, 2020. This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20181354, in part by the Science and Technology Plan of Xi'an under Grant 20191122015KYPT011JC013, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX19_2142, in part by the Postgraduate Research and Practice Innovation Program of China University of Mining and Technology under Grant KYCX19_2142, in part by the National Natural Science Foundation of China under Grant 61771473, Grant 61991451, and Grant 61379143, in part by the Six Talent Peaks High-Level Talents in Jiangsu Province under Grant XYDXX-063, and in part by the Qing Lan Project. This article was recommended by Associate Editor S. Chen. *(Corresponding author: Leida Li.)*

Hancheng Zhu is with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: zhuhancheng@cumt.edu.cn).

Leida Li, Jinjian Wu, and Guangming Shi are with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: ldli@xidian.edu.cn; jinjian.wu@mail.xidian.edu.cn; gmshi@xidian.edu.cn).

Sicheng Zhao is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94710 USA (e-mail: schzhao@gmail.com).

Guiguang Ding is with the School of Software, Tsinghua University, Beijing 100084, China (e-mail: dinggg@tsinghua.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2020.2984670

## I. Introduction

THE INBORN capability of perceiving visual aesthetics is an important aspect of human intelligence. With the development of artificial intelligence, we expect that machines can imitate human beings to automatically evaluate the aesthetics of images. Consequently, image aesthetic assessment (IAA) has become an important research topic due to its widespread applications [1], such as image recommendation [2], [3]; personalized photo album management [4]; perceptual image enhancement [5], [6]; and image retrieval [7], [8]. While human beings can effortlessly gauge the visual aesthetics of images, it remains a great challenge for machines these days.

Existing work for IAA mainly focuses on generic IAA (GIAA) [9]–[17], which infers the common rules of visual aesthetics perceived by an average user [18], typically as binary classification [9], [10] or quality prediction [12], [13]. However, it is well known that human cognitive processes are different among individuals [19]. There is no such thing as average user in real life [18] and the aesthetic perception of image are highly subjective [20]. Individual users may have different aesthetic preferences on the same image, depending on an individual's unique personality traits [21]–[23]; emotions [24], [25]; and so on. As an example, Fig. 1 shows two images and the associated aesthetic scores rated by five individual users from the FLICKR-AES database [4]. For comparison, the average aesthetic scores are also shown. As illustrated in Fig. 1, the average aesthetic score of an image counteracts individual user's aesthetic preferences, so it is difficult to infer the aesthetic perception of an individual user from the average aesthetics. While generic image aesthetics has been extensively researched, to learn individual user's unique aesthetic preference is still an open problem [20].

The problem of learning individual user's visual aesthetic preference is called personalized IAA (PIAA) [4], [26]–[30]. PIAA is a very challenging task since it is difficult to collect a large amount of annotated images for a specific user, which are needed to train an effective prediction model. Therefore, PIAA is a typical small sample learning (SSL) problem [31], which cannot be modeled directly with conventional deep networks. The key step for SSL is to find reliable prior knowledge learned from known tasks that can be exploited for faster learning over unseen tasks. To address the problem of PIAA, several

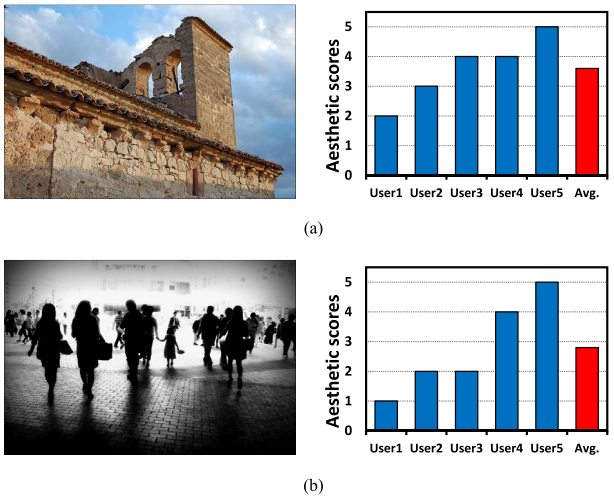


Fig. 1. Example images and the corresponding aesthetic scores rated by five individual users from the FLICKR-AES database [4]. The aesthetic scores range from 1 to 5, and the higher score indicates the higher aesthetics. The blue bar indicates the aesthetic scores rated by five users, and the red bar is the average aesthetic score.

efforts have been carried out recently [4], [28]–[30]. These approaches learn an individual's personalized aesthetic assessment by transferring and adapting the pretrained GIAA model as prior knowledge. However, they suffer from the following limitations.

1) The GIAA model learned from the average aesthetics cannot accurately capture the shared aesthetic prior knowledge when people gauge image aesthetics, since it simply uses the average score as the training target, which counteracts the differences of individual aesthetic perception.
2) The learned GIAA model cannot quickly adapt to unseen PIAA tasks with only a small number of training samples.

In view of the aforementioned problems, an effective approach is to learn an aesthetic prior model with broadly accepted judgments of multiple people [20]. The prior model can fast adapt to a new PIAA task with a few labeled samples [32], [33]. To this end, extensive individual users' visual aesthetic ratings should be directly used to learn the aesthetic's prior model. Particularly, each individual user's aesthetic ratings on images can be considered as a PIAA task, and we can leverage the few-shot learning (FSL) [34] strategy to learn the aesthetic prior model. Furthermore, optimization-based meta-learning [35] is an effective method to refine model parameters learned from extensive tasks. Hence, we utilize an optimization-based meta-learning approach to learn a generalized aesthetic prior model that can quickly adapt to unseen PIAA tasks.

In this article, we introduce a novel personalized aesthetics assessment method based on meta-learning. The proposed approach leverages the bilevel gradient descent strategy from extensive PIAA tasks to learn an aesthetic prior model, which can quickly adapt to a new PIAA task using a small amount of training samples. The contributions of this article are summarized as follows.

1) We explore the problem of IAA from the perspective of individualization. Considering the small sample property of personalized IAA, we treat the aesthetic evaluation of each user as an independent PIAA task and leverage a meta-learning approach to learn the prior knowledge from extensive PIAA tasks of different users in visual aesthetic appreciation. The prior knowledge learns the shared rule of how people judge image aesthetics.
2) We propose a personalized aesthetic assessment approach with strong expansibility, which can be applied to any deep regression networks. This approach can learn the shared aesthetic judgment of different people by refining model parameters with explicit gradient optimization from numerous PIAA tasks.
3) We propose using bilevel gradient optimization to learn a prior model that has the ability of fast adaptation for individual users' aesthetic preferences. The proposed PIAA method has better generalization performance than the state-of-the-art methods on several databases.

The remainder of this article is structured as follows. The related works of PIAA and meta-learning are briefly introduced in Section II. In Section III, the proposed meta-learning-based PIAA approach is presented together with the bilevel gradient optimization strategy. Extensive experimental results and comparisons are given in Section IV, and finally, the conclusions are drawn in Section V.

## II. Related Work

In this article, we address the problem of PIAA, which is based on meta-learning. In this section, we give a brief description of the earlier works related to these two parts.

### A. Personalized Image Aesthetics Assessment

Most studies of IAA hold that there are generic rules for human visual aesthetics [1]. Consequently, IAA is usually considered as a binary classification [9]–[11] or regression task [12], [13] to predict high- and low-aesthetic categories, or to predict an aesthetic score. For example, Tang *et al.* [9] proposed extracting multiple regional features in different ways according to image content for binary aesthetics classification. Kong *et al.* [12] proposed learning a ranking model based on the Siamese network [36] for image aesthetic score regression. Realizing that people's aesthetic preferences on the same image may be different, several works have demonstrated that predicting image aesthetic distribution is more effective in describing the diversified aesthetics of images [14]–[17]. Although the aesthetics distribution prediction can reflect the diversity of human visual aesthetics to a certain extent, it is still a coarse statistics of people's aesthetic preferences. IAA needs to be refined to an individual user level for better personalized customization [4], [26]–[30]. Lv *et al.* [26] leveraged the ranking model and user interaction to learn users' personal aesthetic preferences. This approach requires user's real-time participation. Wang *et al.* [27] proposed a collaborative filtering-based approach with user-image textual reviews for PIAA. This method assumes that there is a considerable

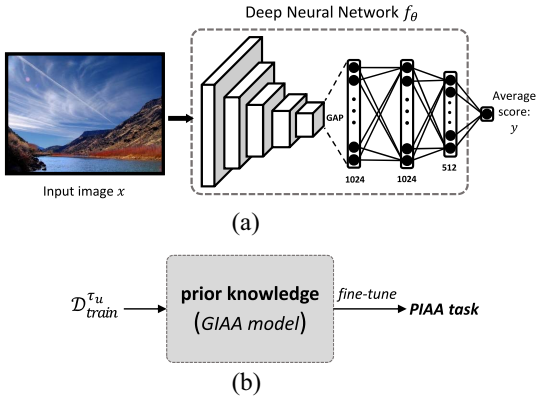


Fig. 2. Illustration of the PIAA approach using the GIAA model based on average aesthetics as prior knowledge. (a) Network structure of the GIAA model. (b) Fine-tune the GIAA model to PIAA task.

overlap between the images evaluated by different users, which may not hold firmly in some applications.

To deal with the SSL problem of PIAA, the approaches in [4] and [28]–[30] take the GIAA model learned from the average aesthetic scores of images as prior knowledge and fine-tuned the GIAA model for the user-specific PIAA model. Li *et al.* [28] proposed a personality-assisted multi-task deep-learning framework that can handle both generic and personalized IAA. This approach uses the GIAA model as a prior model, and further leverages the relationship between users' personality traits and image preferences to predict users' personalized image aesthetics. Fig. 2 shows a typical PIAA pipeline using the average aesthetics-based GIAA model as prior knowledge. In this pipeline, images and the corresponding average scores are used to train a GIAA model and the GIAA model is fine-tuned on a few user-specific training samples for the PIAA task. However, the average score of an image can hardly embody the aesthetic differences of users, which makes the GIAA model difficult to adapt to individual users' aesthetic preferences quickly, so it is not reliable prior knowledge for individual aesthetic perception. Therefore, it is crucial to acquire shared prior knowledge, namely, the shared rule that people judge image aesthetics, from different individual users in the visual aesthetic experience. In contrast to the existing PIAA approaches, we resort to meta-learning for learning the shared aesthetic prior knowledge from extensive PIAA tasks.

### *B. Meta-Learning*

Meta-learning is a machine-learning technique [32] to solve the problem of learning how to learn. Despite the data-driven deep-learning technique succeeds in many specific tasks (e.g., image classification [37], [38] and object detection [39], [40]), it still lacks the ability to learn from limited samples and quickly generalize to new tasks. To address this problem, meta-learning imitates human's ability to acquire prior knowledge from extensive learning tasks, which can be quickly adapted to new tasks. FSL [34] is a typical meta-learning problem that aims to train an effective learning model from very few samples. There are three main approaches: 1) recurrent neural-networks (RNNs) memory-based methods [41], [42]; 2) metric-based methods [43], [44]; and 3) optimization-based methods [35], [45]. The RNN memory-based methods use RNNs with memories to store experience knowledge from the previous tasks for learning new task [41], [42]. The metric-based methods mainly learn an embedding function that maps the input space to a new embedding space, and leverage nearest neighbor or linear classifiers for image classification [43], [44]. The optimization-based methods aim to learn the initialization parameters of a model that can quickly learn new tasks by fine-tuning the model using few training samples [35], [45]. The essential purpose of these approaches is to learn prior knowledge among a wide range of learning tasks for quickly adapting to unseen tasks.

This article is related to the optimization-based meta-learning methods, which try to learn a model that can rapidly adapt to new tasks from the previous tasks. Finn *et al.* [35] proposed a model-agnostic meta-learning (MAML) approach to optimize model parameters by updating second-order stochastic gradient descent (SGD). The computational complexity of the second-order gradient is too high for large-scale network training. Therefore, a first-order gradient-based meta-learning method called Reptile was proposed in [45]. Franceschi *et al.* [46] leveraged a bilevel optimization framework [47] to unify gradient-based optimization and meta-learning. In this framework, the lower level optimization denotes the adaptation to a given task and the upper level optimization represents the training of meta-learner. In this way, the upper level optimization enables the meta-learner to learn the cross-task-related information and knowledge on similar tasks, and the lower level optimization can capture the specific information on different tasks, making the model more adaptive for new tasks [48]. However, these works cannot be directly applied in the proposed PIAA task, because they are primarily designed for FSL in the classification task, where the number of training samples in each task is typically less than 10 [34]. In contrast, the PIAA task requires a continuous measure of image aesthetic quality, which makes the PIAA problem different and more complex.

In view of the PIAA task, we adopt a bilevel gradient optimization method that is different from the previous approaches. The existing bilevel optimization methods usually use the upper level optimization to jointly update the model parameters of different tasks in the lower level optimization [48]. In contrast, our PIAA approach needs to further learn the fast adaptability of model parameters of the same task in lower level optimization. The difference of the bilevel optimization between our method and the previous approaches can be summarized as follows. First, we divide the annotated data in each user's PIAA task into two sets for two-level updating to learn the fast adaptability from one set to another in a task. Second, we introduce the Adam optimizer [49] into the meta-training stage of our aesthetic metamodel. Since our aim is to learn the prior model with the shared aesthetic judgment of different users, we then learn the shared information and knowledge among different PIAA tasks during the upper level optimization. Therefore, a bilevel gradient optimization-based meta-learning is introduced to learn a better aesthetic prior model for the PIAA task.

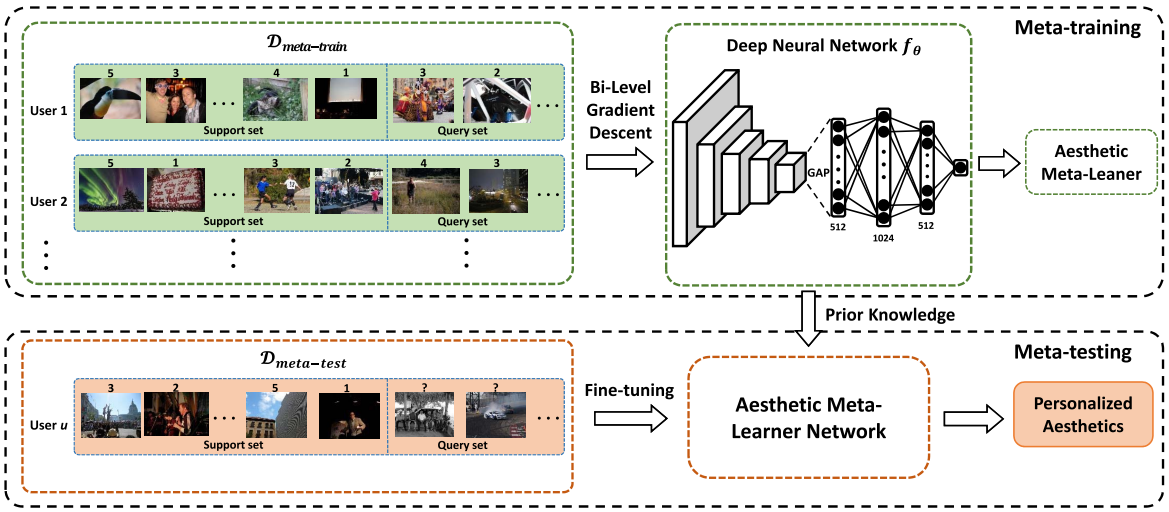


Fig. 3. Framework of the proposed PIAA method based on meta-learning with bilevel gradient optimization.

## III. Proposed Method

In this section, we propose a PIAA method based on meta-learning with bilevel gradient optimization. This article uses meta-learning to seek an aesthetic prior model with people's aesthetic judgment from extensive PIAA tasks. In other words, the prior model tries to learn the shared rule that different people judge image aesthetics. The aesthetic prior model is called aesthetic meta-learner. Fig. 3 shows the framework of the proposed approach, which consists of two phases: 1) meta-training and 2) meta-testing. In meta-training, we first leverage a large number of individual users' PIAA tasks to produce a meta-training set, which is further divided into two sets: 1) support set and 2) query set. Then, a bilevel gradient descent method is employed to train an aesthetic meta-learner model using the support set and query set. In meta-testing, the support set of an individual user is used to fine-tune the aesthetic meta-learner network for obtaining the PIAA model. The proposed approach is called bilevel gradient optimization-based PIAA (BLG-PIAA).

### A. Meta-Training Phase

*1) Meta-Learning for Shared Aesthetic Prior Knowledge:* As illustrated in Fig. 1, the prior model based on average aesthetics may be problematic for PIAA tasks. Consequently, the most critical issue is how to learn an aesthetic model that can extract shared prior knowledge from individual aesthetic data directly during training. In contrast to the previous PIAA approaches [4], [28]–[30], we treat the personalized aesthetics assessment of each user as a task, producing a large number of PIAA tasks. Inspired by the idea of meta-learning for learning to learn [32], we leverage an optimization-based meta-learning approach to refine model parameters from extensive users' PIAA tasks. For the PIAA task, it is important that the learned aesthetic prior model can be easily fine-tuned with a small number of training samples. Therefore, we use a bilevel optimization method to learn the shared prior knowledge across different PIAA tasks. In order to learn the ability of fast adaptation in each PIAA task, the training data of each PIAA task are further divided into the support set and query set. The support set is first used to tentatively update the model parameters, and then the updated model validates whether it is well performed in the query set. The two-level gradient updating approach from the support set to query set is called bilevel gradient optimization.

*2) Bilevel Gradient Optimization:* We adopt a deep convolutional neural network (CNN) that is pretrained on the ImageNet dataset [37] as our basic backbone and remove the fully connected layers after the last convolutional layer. Then, a global average pooling (GAP) operation and two-layer fully connected neural network are employed as the fully connected (FC) layers of our deep neural network $f_\theta$. Following each FC layer, dropout and batch normalization (BN) are added to control overfitting and accelerate the convergence rate of training. In particular, for an input image $x$, we fed it into the deep network to generate the predicted aesthetic score $\hat{y}$, which is defined as

$$\hat{y} = f_\theta(x; \theta) \tag{1}$$

where $\theta$ denotes the initialized network parameters. Since we expect to minimize the difference between the predicted and ground-truth aesthetic scores of the image $x$, the squared Euclidean distance is used as the loss function, which takes the following form:

$$\mathcal{L} = \|f_\theta(x; \theta) - y\|_2^2 \tag{2}$$

where $y$ denotes the ground-truth aesthetic score of the input image $x$.

In our PIAA task, we first collect $m$ annotated samples from each user and divide them into a support set and a query set, where the support set contains $m_s$ samples and the query set contains $m_q$ samples ($m = m_s + m_q$). Then, we denote $\mathcal{D}^{p(\tau)}_{\text{meta-train}} = \{(\mathcal{D}^{\tau_i}_{tr_s}, \mathcal{D}^{\tau_i}_{tr_q})\}_{i=1}^{N}$ as the meta-training set of PIAA tasks, where $\mathcal{D}^{\tau_i}_{tr_s}$ and $\mathcal{D}^{\tau_i}_{tr_q}$ are the $i$th support set and query set of PIAA task, and $N$ is the number of PIAA tasks. We sample $k$ PIAA tasks as a batch from the meta-training set ($1 < k \leq N$). For the $i$th support set $\mathcal{D}^{\tau_i}_{tr_s}$ in the

batch, the loss can be calculated by (2) and denoted as $\mathcal{L}_{\tau_i}$ ($i \in \{1, 2, \ldots, k\}$). As far as we know, the SGD method cannot dynamically update the learning rate and easily fall into the local optimal solution [50]. In order to adaptively learn the gradient agreement between each step in a task, we leverage an efficient gradient descent method to optimize the proposed model. Therefore, we first calculate the first-order gradients of loss function $\mathcal{L}_{\tau_i}$ relating to all model parameters and it is defined as

$$g_\theta = \nabla_\theta \mathcal{L}_{\tau_i}(f_\theta). \tag{3}$$

Then, we use the Adam [49] optimizer to update the model parameters for $S$ steps on the support set $\mathcal{D}_{tr_s}^{\tau_i}$, which is formulated as

$$\text{Adam}\big(\mathcal{L}_{\tau_i}, \theta\big) : \theta_i' \leftarrow \theta - \alpha \sum_{j=1}^{S} \frac{m_{\theta^{(j)}}}{\sqrt{v_{\theta^{(j)}}} + \epsilon} \tag{4}$$

where $\epsilon = 1e-8$ and $\alpha$ is the inner learning rate. $m_{\theta^{(j)}}$ and $v_{\theta^{(j)}}$ are the first moment and second raw moment of gradients, which are defined as

$$m_{\theta^{(j)}} = \mu_1 m_{\theta^{(j-1)}} + (1-\mu_1) g_{\theta^{(j)}} \tag{5}$$

$$v_{\theta^{(j)}} = \mu_2 v_{\theta^{(j-1)}} + (1-\mu_2) g_{\theta^{(j)}}^2 \tag{6}$$

where $m_{\theta^{(0)}} = 0$ and $v_{\theta^{(0)}} = 0$. $\mu_1$ and $\mu_2$ are the exponential decay rates of $m_{\theta^{(j)}}$ and $v_{\theta^{(j)}}$, respectively. $g_{\theta^{(j)}}$ denotes the updated gradients in step $j$ ($j \in \{1, 2, \ldots, S\}$). As mentioned previously, we expect that the prior model updated with the support set can perform well on the query set. Different from [35], we then compute the first-order gradient and update the model parameters a second time. The model parameters $\theta_i'$ are updated with the Adam optimizer for $S$ steps on the query set $\mathcal{D}_{tr_q}^{\tau_i} (i = 1, 2, \ldots, k)$, which takes the form

$$\text{Adam}\big(\mathcal{L}_{\tau_i}, \theta_i'\big) : \theta_i \leftarrow \theta_i' - \alpha \sum_{j=1}^{S} \frac{m_{\theta'^{(j)}}}{\sqrt{v_{\theta'^{(j)}}} + \epsilon} \tag{7}$$

where $m_{\theta'^{(j)}}$ and $v_{\theta'^{(j)}}$ are the first moment and second raw moment of gradients $g_{\theta'^{(j)}}$, which denotes the updated gradients in step $j$ ($j \in \{1, 2, \ldots, S\}$). For a batch of $k$ PIAA tasks, we calculate the gradient agreement of all tasks to update the model parameters, which is defined as

$$\theta \leftarrow \theta - \beta \frac{1}{k} \sum_{i=1}^{k} (\theta - \theta_i) \tag{8}$$

where $\beta$ is the outer learning rate. In this way, $k$ PIAA tasks on the meta-training set $\mathcal{D}_{\text{meta-train}}^{p(\tau)}$ are sampled iteratively for model training. Finally, an aesthetic meta-learner model can be obtained.

### B. Meta-Testing Phase

After training the aesthetic meta-learner model from extensive PIAA tasks, we then use this model as prior knowledge for fine-tuning. Given $M$ images rated by a testing user, we denote the support set as $\mathcal{D}_{te_s}^{\tau_u} = \{x_j, y_j\}_{j=1}^{M}$. We first use the squared Euclidean distance as the loss function and calculate the gradients of model parameters, which are formulated as

$$\mathcal{L}_{\tau_u} = \frac{1}{M} \sum_{j=1}^{M} \big\| \hat{y}_j - y_j \big\|_2^2 \tag{9}$$

$$g_\theta = \nabla_\theta \mathcal{L}_{\tau_u}(f_\theta) \tag{10}$$

where $\hat{y}_j$ is the predicted personalized score of the $j$th image. Then, we use the Adam optimizer to update the prior model parameters for $P$ epochs on the support set $\mathcal{D}_{te_s}^{\tau_u}$, which is formulated as

$$\text{Adam}\big(\mathcal{L}_{\tau_u}, \theta\big) : \theta_{te} \leftarrow \theta - \alpha_f \sum_{j=1}^{P} \frac{m_{\theta^{(j)}}}{\sqrt{v_{\theta^{(j)}}} + \epsilon} \tag{11}$$

where $\alpha_f$ is the learning rate of fine-tuning. $m_{\theta^{(j)}}$ and $v_{\theta^{(j)}}$ are first moment and second raw moment of gradients $g_{\theta^{(j)}}$ ($j \in \{1, 2, \ldots, P\}$). Finally, the user-specific PIAA model is obtained for predicting personalized aesthetics of images. It is worth noting that the training process of our personalized model does not need to learn additional parameters other than the deep network $f_\theta$, which greatly improves the learning efficiency.

For a query image $x$, we fed it into the PIAA model $f_{\theta_{te}}$ to generate the predicted personalized aesthetics score $\hat{y} = f_{\theta_{te}}(x; \theta_{te})$. The whole procedure of our algorithm is outlined in Algorithm 1.

**Algorithm 1** Bilevel Gradient Optimization for PIAA

**Input:** Meta-training set $\mathcal{D}_{meta-train}^{p(\tau)} = \{(\mathcal{D}_{tr_s}^{\tau_i}, \mathcal{D}_{tr_q}^{\tau_i})\}_{i=1}^{N}$, where $\mathcal{D}_{tr_q}^{\tau_i}$ and $\mathcal{D}_{tr_s}^{\tau_i}$ are the $i$th support set and query set of PIAA task, and $N$ is the number of PIAA tasks, support set of a testing user's PIAA task $\mathcal{D}_{te_s}^{\tau_u}$, query image $x$, outer learning rate $\beta$

**Output:** Predicted personalized aesthetic score $\hat{y}$ for $x$

```
Initialize model parameters θ: pre-trained on Imagenet;
/* meta-training phase */
for iteration = 1, 2, ... do
    Sample a batch of k tasks in D_{meta-train}^{p(τ)};
    for i = 1, 2, ..., k do
        /* first level computing */
        Compute θ_i' = Adam(L_{τ_i}, θ) for S steps on D_{tr_s}^{τ_i};
        /* second level computing */
        Compute θ_i = Adam(L_{τ_i}, θ_i') for S steps on D_{tr_q}^{τ_i};
    end for
    update θ ← θ − β (1/k) Σ_{i=1}^{k} (θ − θ_i);
end for
/* meta-testing phase */
Update θ_te = Adam(L_{τ_u}, θ) for P epochs on D_{te_s}^{τ_u};
Input x into the updated model f_{θ_te};
return ŷ.
```

## IV. Experimental Results and Discussion

### A. Experimental Settings

*1) Databases:* We evaluate the effectiveness of the proposed approach for PIAA on two large-scale databases

(i.e., FLICKR-AES [4] and AADB [12]) and one small-scale database (i.e., REAL-CUR [4]). Ren *et al.* [4] first introduced FLICKR-AES and REAL-CUR databases with rater's ID for PIAA task. Besides, the AADB database [12] also provides rater's ID when labeling the aesthetic scores of images.

The *FLICKR-AES database* [4] includes 40 000 images downloaded from Flickr, which are rated by a total of 210 workers through AMT. The aesthetic scores range from 1 to 5, and higher score indicates higher aesthetics. In this database, 35 263 images rated by 173 workers constitute the training set, and the remaining 4737 images labeled by 37 workers constitute the testing set. The number of images annotated by each testing worker ranges from 105 to 171. For PIAA tasks, the training set is used for learning an aesthetics model with prior knowledge, and the testing set is used for learning the personalized aesthetics model of each testing worker.

The *AADB database* [12] includes about 10 000 images, which are rated by a total of 190 workers. For each image, five workers rated and provided their annotated aesthetic scores and 11 aesthetic attributes. The 11 attributes are *interesting content, object emphasis, good lighting, color harmony, vivid color, shallow depth of field, motion blur, rule of thirds, balancing element, repetition*, and *symmetry*. The aesthetic scores range from 1 to 5, and the higher score indicates higher aesthetics. The AADB database is originally used for evaluating the performance of GIAA methods. In order to evaluate the performance of our PIAA method on the AADB database, we use 22 workers and their rated images as the testing set, and the remaining 168 workers and the labeled images as the training set. The number of images annotated by each testing worker ranges from 110 to 190. We leverage images included in the training set for learning a prior model and fine-tune it on the testing set for PIAA tasks.

*REAL-CUR* [4] includes 14 users' personal albums and the corresponding aesthetic scores on their own photo albums. The aesthetic scores range from 1 to 5. The number of images in each user's album ranges from 197 to 222. This small-scale database can be used to verify the effectiveness of the aesthetic prior model learned from the training set of large-scale databases (e.g., FLICKR-AES and AADB) for PIAA tasks in real-world applications.

*2) Implementation Details:* Three popular deep-learning architectures, that is, AlexNet [37], ResNet18 [51], and Inception-v3 [52], are adopted as our basic CNN layers, which are pretrained on ImageNet [37]. The two fully connected layers are randomly initialized. In the proposed approach, the inner learning rate $\alpha$ of basic CNN layers and fully connected layers is set as $1e-4$ and $1e-3$, respectively, and the outer learning rate $\beta$ is set as $1e-2$. The learning rate of fine-tuning $\alpha_f$ is set to $1e-5$. The learning rates drop to a factor of 0.9 after every 100 iterations. In the meta-training phase, the number of samples in support set $m_s$ and query set $m_q$ of each task is set to 80 and 20, respectively. The number of tasks $k$ in a batch is 4. The step size of learning $S$ is 5 and the epoch of fine-tuning $P$ is 20. The total iteration of prior model training is set to 1000. The weight decay is $1e-5$. The exponential decay rates $\mu_1$ and $\mu_2$ are set as 0.9 and 0.99. The proposed method is implemented based on PyTorch [53]. The source code of our proposed method is publicly available.[1]

*3) Baseline Methods:* To evaluate the effectiveness of our PIAA method based on bilevel gradient optimization (BLG-PIAA), we compare our approach with three state-of-the-art methods (i.e., FPMF [54], PAM [4], USAR [26], and PA_IAA [28]). To further demonstrate the effectiveness of our approach with bilevel gradient optimization, we further implement the proposed model using two other optimization methods [i.e., PIAA(MAML) and PIAA(Reptile)] and a baseline PIAA method (BA-PIAA) using the GIAA model as prior knowledge, which have the same network structure as our BLG-PIAA.

1) *FPMF* [54] is a collaborative filtering approach to recommend color aesthetics for individual user and the testing results in three forms [FPMF (only attribute), FPMF (only content), and FPMF (content and attribute)] on the FLICKR-AES database are released in [4].
2) *PAM* [4] is an active personalized aesthetics model that leverages image attributes and contents to predict the residual of individual user's aesthetic ratings. Three forms of this approach, PAM (only attribute), PAM (only content), and PAM (content and attribute), are tested on the FLICKR-AES database.
3) *USAR* [26] is a personalized image aesthetic ranking approach by incorporating individual user's interaction for PIAA. Three methods of user's interaction, including USAR_PPR, USAR_PAD, and USAR_PPR&PAD, have been tested on the FLICKR-AES database.
4) *PA_IAA* [28] is a personality-assisted multitask deep-learning method that takes advantage of people's Big-Five personality traits to predict individual users' aesthetic preferences on images.
5) *PIAA(MAML) and PIAA(Reptile)* are two versions of the proposed model by substituting our bilevel gradient optimization with MAML [35] and Reptile [45], where the SGD operation is employed to learn the aesthetic prior model from extensive users' PIAA tasks and fine-tuning is conducted on the aesthetic data of a target user to obtain the PIAA model.
6) *BA-PIAA* is a baseline PIAA method based on average aesthetics (see Fig. 2). In this approach, we first leverage images with the corresponding average scores to train a GIAA model and then the GIAA model is fine-tuned on a small number of training data for PIAA task. During the training and fine-tuning processes, the Euclidean loss function and the Adam optimizer are also used to optimize the BA-PIAA model.

*4) Evaluation Criteria:* For PIAA task, the ranking consistency between predicted and ground-truth results is a very important evaluation criteria [4], [26], [28]. We employ the Spearman rank-order correlation coefficient (SROCC) [55] to evaluate the performance of PIAA approaches. Supposing $s_i$ and $\hat{s}_i$ denote the ranks of the $i$th testing image in ground truth and predicted aesthetic scores, respectively, the difference between the ground truth and predicted aesthetic scores

[1] https://github.com/zhuhancheng/BLG-PIAA

TABLE I
COMPARISON RESULTS (SROCC) OF BA-PIAA, BLG-PIAA, AND THE STATE-OF-THE-ART METHODS ON FLICKR-AES

| Method | 10 images | 100 images |
|---|---|---|
| FPMF (only attribute) [54] | 0.511±0.004 | 0.516±0.003 |
| FPMF (only content) [54] | 0.512±0.002 | 0.516±0.010 |
| FPMF (content and attribute) [54] | 0.513±0.003 | 0.524±0.007 |
| PAM (only attribute) [4] | 0.518±0.003 | 0.539±0.013 |
| PAM (only content) [4] | 0.515±0.004 | 0.535±0.017 |
| PAM (content and attribute) [4] | 0.520±0.003 | 0.553±0.012 |
| 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 |
| PA_IAA [28] | 0.543±0.003 | 0.639±0.011 |
| PIAA(MAML) | 0.520±0.005 | 0.569±0.010 |
| PIAA(Reptile) | 0.529±0.006 | 0.598±0.015 |
| BA-PIAA | 0.524±0.004 | 0.583±0.014 |
| **BLG-PIAA** | **0.561±0.005** | **0.669±0.013** |

is computed as $d_i = s_i - \hat{s}_i$, and the SROCC is defined as

$$\text{SROCC} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \tag{12}$$

where $N$ is the number of the testing images. If there is a perfect positive (negative) correlation between $s_i$ and $\hat{s}_i$ ($i = 1, 2, \ldots, N$), the SROCC equals 1 ($-1$). Therefore, the SROCC ranges from $-1$ to 1, and higher absolute SROCC value indicates better performance.

### *B. Performance on the FLICKR-AES Database*

Similar to [4], [26], and [28], we also evaluate our approach on the testing workers on the FLICKR-AES database. For each worker, the images he or she labeled are randomly divided into two sets, that is, $M$ images on the support set for training and the remaining images on the query set for testing. The experiments are conducted 50 times for each worker to avoid the bias of randomness, and the averaged results as well as the standard deviation are reported. To compare with the reported results of existing PIAA methods in [4], [26], and [28], we also set $M = 10$ and $M = 100$, respectively.

Table I lists the testing results of BA-PIAA, PIAA(MAML), PIAA(Reptile), BLG-PIAA, and the state-of-the-art methods for PIAA, and the best results are highlighted in bold font. In our implementation, ResNet18 is used as the basic backbone. We can see that BLG-PIAA achieves higher SROCC values than the state-of-the-art approaches based on collaborative filtering (FPMF [54]), user's interaction (USAR_PPR&PAD [26]), and generic prior aesthetics (PAM [4] and PA_IAA [28]). Although the proposed BA-PIAA based on average aesthetics yields very encouraging performance compared with these PIAA methods, BLG-PIAA further achieves 3.7% and 8.6% performance improvement when 10 and 100 images are used for training, respectively. Besides, our BLG-PIAA model is also significantly better than the other optimization-based approaches [PIAA(MAML) and PIAA(Reptile)]. This demonstrates that our meta-learning approach based on bilevel gradient optimization is very effective for the PIAA task.

TABLE II
COMPARISON RESULTS (SROCC) OF THE PROPOSED BA-PIAA AND BLG-PIAA BASED ON THREE BASIC BACKBONES (ALEXNET, RESNET18, AND INCEPTION-V3) ON FLICKR-AES

| Basic backbone | Method | 10 images | 100 images |
|---|---|---|---|
| AlexNet | BA-PIAA | 0.491±0.002 | 0.556±0.007 |
| | BLG-PIAA | 0.534±0.003 | 0.624±0.011 |
| ResNet18 | BA-PIAA | 0.524±0.004 | 0.583±0.006 |
| | BLG-PIAA | 0.561±0.005 | 0.669±0.013 |
| Inception-v3 | BA-PIAA | 0.519±0.004 | 0.576±0.009 |
| | BLG-PIAA | 0.548±0.006 | 0.651±0.016 |

For each testing worker, the ability of effectively learning personalized aesthetic preferences from the aesthetic prior model is quite important. To verify the generalization ability of the aesthetic prior model, we first directly use the prior model of BA-PIAA and BLG-PIAA to compute the prediction performance of 37 workers, and then further calculate the testing results of each worker by using the worker-specific personalized model of BA-PIAA and BLG-PIAA when $M = 100$. The prediction performances (SROCC) are shown in Fig. 4. We can find that the prediction performance of the personalized model outperforms a prior model for both BA-PIAA and BLG-PIAA. Particularly, the average increases of SROCC values for 37 workers tested on the BA-PIAA model and the BLG-PIAA model are about 0.079 (from 0.504 to 0.583) and 0.129 (from 0.540 to 0.669), respectively. Compared with BA-PIAA, BLG-PIAA not only obtains a more effective prior model (0.540 versus 0.504) but also has better performance in fast learning personalized aesthetic preferences from the prior model (0.129 versus 0.079).

To further demonstrate the effectiveness of our BLG-PIAA model using different network architectures, we compare BLG-PIAA with BA-PIAA based on three popular backbones (AlexNet, ResNet18, and Inception-v3) on the FLICKR-AES database. The testing results are summarized in Table II. As expected, the proposed BLG-PIAA outperforms BA-PIAA by a large margin regardless of basic backbone. It is worth noting that BLG-PIAA and BA-PIAA have the same network parameters for each basic backbone. Compared with BA-PIAA, BLG-PIAA is a more efficient and extensible personalized aesthetic assessment method that can improve the performance without changing the network structure. Furthermore, the proposed BLG-PIAA methods based on three basic backbones all outperform the state-of-the-art PIAA approaches listed in Table I. This further demonstrates the advantage of the proposed aesthetic prior model.

### *C. Performance on AADB and REAL-CUR Databases*

As far as we know, no available PIAA method has released the testing results on AADB and REAL-CUR. To verify the performance of the proposed method on AADB and REAL-CUR databases, we compare BLG-PIAA with BA-PIAA based on ResNet18. For AADB, we use the images on the training set to train a prior model and randomly select $M$ images rated by 22 testing workers to fine-tune the prior model. For the small-scale REAL-CUR database, we leverage the model

TABLE III
COMPARISON RESULTS (SROCC) OF THE PROPOSED BA-PIAA AND BLG-PIAA BASED ON RESNET18 ON AADB AND REAL-CUR DATABASES

| Database | Method | 10 images | 100 images |
|---|---|---|---|
| AADB | BA-PIAA | 0.450±0.001 | 0.513±0.005 |
| | BLG-PIAA | 0.497±0.003 | 0.545±0.007 |
| REAL-CUR | BA-PIAA | 0.407±0.005 | 0.534±0.018 |
| | BLG-PIAA | 0.448±0.007 | 0.578±0.015 |

TABLE IV
COMPARISON RESULTS (SROCC) OF BLG-PIAA BASED ON RESNET18 WHEN TRAINING AND TESTING ON DIFFERENT DATABASES

| Train Database | Test Database | | |
|---|---|---|---|
| | FLICKR-AES | AADB | REAL-CUR |
| FLICKR-AES | 0.669 | 0.526 | 0.578 |
| AADB | 0.601 | 0.545 | 0.538 |

learned from the training set on FLICKR-AES as a prior model and fine-tune it with $M$ randomly selected images from each album on REAL-CUR. In this experiment, we set $M = 10$ and $M = 100$, and the averaged results and the standard deviation of 50 times are reported in Table III. As can be seen, the performance of BLG-PIAA is significantly better than those of BA-PIAA on AADB and REAL-CUR databases. In particular, the proposed method based on meta-learning is more effective than the one based on average aesthetics in the real-world IAA of personal albums. This demonstrates that our BLG-PIAA model can be quickly generalized to the PIAA tasks of real-world users.

### D. Cross Database Evaluation

In practical applications, we expect the learned aesthetic prior model can quickly adapt to a user's PIAA model by just performing fine-tuning on a small amount of aesthetic data. Therefore, we need to learn the prior model from the training users in one database and verify its generalization ability to other testing users in another database. In order to validate the generalization ability of the learned prior model for the PIAA tasks of unknown users in different databases, we conduct a cross database evaluation in this experiment. Particularly, the BLG-PIAA models based on ResNet18 learned on the training sets of FLICKR-AES and AADB are fine-tuned across all the testing sets of three databases. For each PIAA task on three testing sets, the number of fine-tuned images is 100. We conduct the experiments 50 times and list the averaged results in Table IV. For the large-scale databases FLICKR-AES and AADB, training on one database and testing on the other database yield the performance that is close to the one training and testing on the same database. For example, training on FLICKR-AES and testing on AADB achieve 0.526 ranking correlation, which is close to the performance of training and testing on AADB (0.545). We note that the model trained using AADB has better performance when tested on the FLICKR-AES database. This may be due to the fact that the testing users on FLICKR-AES provide more accurate annotated samples for model fine-tuning than those on AADB. This makes the PIAA tasks of testing users in the AADB database more difficult to predict than in the FLICKR-AES database (0.545 versus 0.601). For the small-scale database REAL-CUR, the BLG-PIAA model trained on FLICKR-AES outperforms the one trained on AADB, because FLICKR-AES contains more training images than AADB, which allows that the BLG-PIAA model trained on FLICKR-AES has better generalization ability.

### E. Parameters Discussion and Analysis

In the meta-training phase of our BLG-PIAA, there are four key parameters, that is, the number of samples $m_s$ and $m_q$ in the support set and query set of each PIAA task and the number of PIAA tasks $k$ for gradient optimization together and $S$ to control the learning steps of each PIAA task. In this experiment, the meta-training set can be generated from the PIAA tasks of 173 training workers on the FLICKR-AES database and the number of samples in each training worker is set to 100 ($m = 100$). Then, we set $m_s$, $m_q$, $k$, and $S$ to different values in the meta-training phase and show the averaged SROCC results of 37 testing workers in Fig. 5. In the meta-testing phase, the number of fine-tuning images of each testing workers is 100 ($M = 100$). Overall, when our metamodel is trained with two-level gradient updating from the support set to query set [Fig. 5(b)–(f)], the performance is significantly better than that of merging the support set and query set together [Fig. 5(a)]. This demonstrates the effectiveness of our bilevel gradient optimization approach for the PIAA task. Furthermore, when $m_s = 80$ and $m_q = 20$ [Fig. 5(c)], our method achieves the best overall prediction performance. As can be seen from Fig. 5(c), with the increase of $k$ and $S$, our BLG-PIAA model achieves better prediction performance. When $S$ increases from 1 to 5, the prediction performance increases dramatically. When $k$ is larger than 4, the prediction performance trends to be stable. The parameter $k$ is the number of PIAA tasks for gradient optimization together. The larger $k$ leads to the larger batch size of model training, which is unrealistic when $k$ is too large. Therefore, we set $m_s = 80$, $m_q = 20$, $k = 4$, and $S = 5$ in our experiments.

### F. Visual Analysis for Personalized Image Aesthetics

In order to visualize the performance of our method in PIAA, we randomly select some example images rated by three testing users from FLICKR-AES, AADB, and REAL-CUR databases and evaluate them with the personalized model of BA-PIAA and BLG-PIAA. The number of training images for model fine-tuning is $M = 100$ and the testing results are shown in Fig. 6. From this figure, we can see that the personalized model of BLG-PIAA predicts the user's aesthetic score more accurately than the personalized model of BA-PIAA. This indicates that BLG-PIAA can learn the aesthetic characteristics of people better from extensive PIAA tasks and accurately adapt to the aesthetic preferences of individual users through a small number of fine-tuning samples.

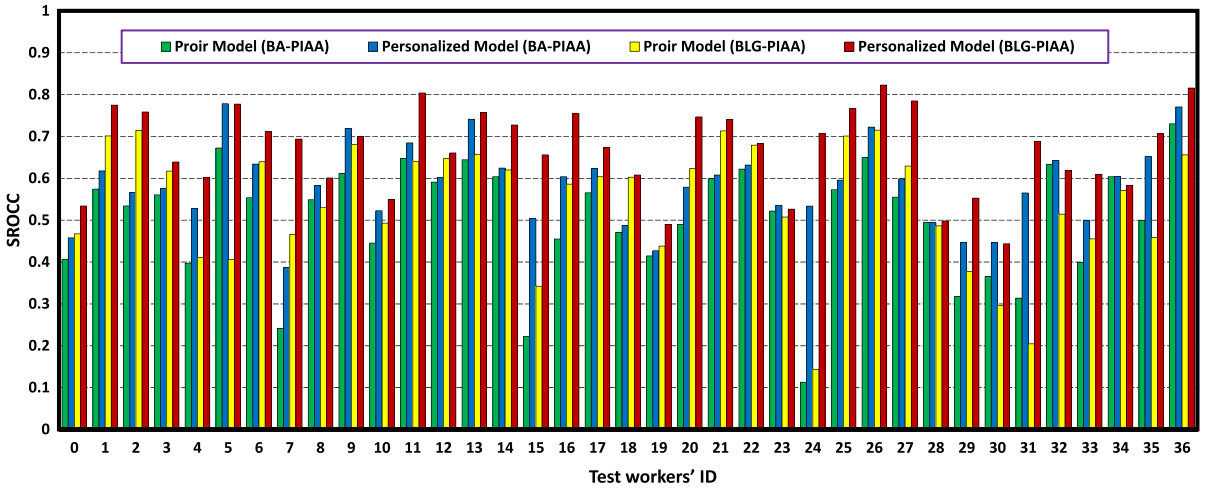


Fig. 4. Performance comparison of 37 testing workers on FLICKR-AES by directly using the aesthetic prior model and aesthetic personalized model when $M = 100$. The green and yellow bars show the SROCC values using the prior model of BA-PIAA and BLG-PIAA, and the blue and red bars show the SROCC values using the personalized model of BA-PIAA and BLG-PIAA.

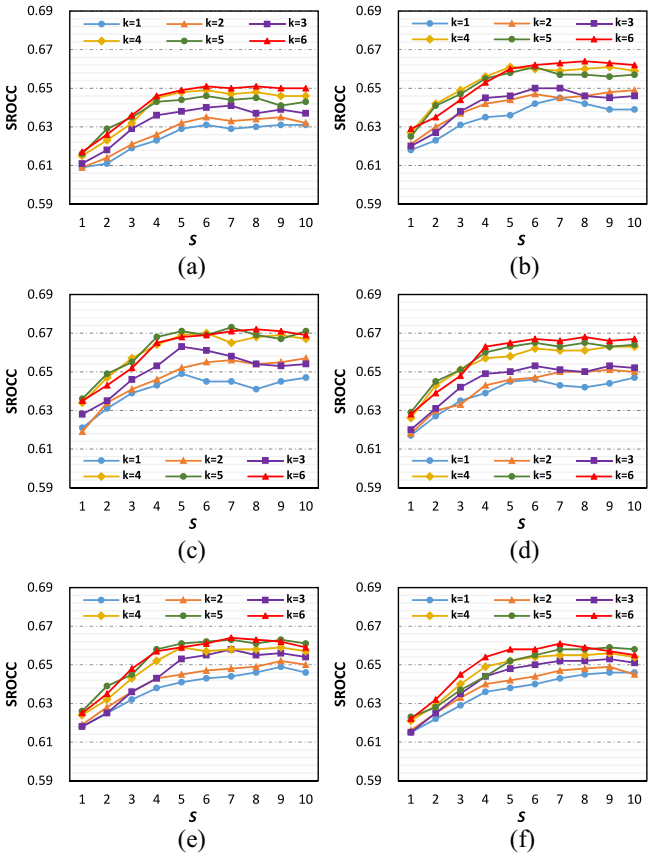


Fig. 5. Influence of parameters $m_s$, $m_q$, $k$, and $S$ in the meta-training of our BLG-PIAA on FLICKR-AES measured by SROCC. (a) $m_s = 100; m_q = 0$. (b) $m_s = 90; m_q = 10$. (c) $m_s = 80; m_q = 20$. (d) $m_s = 70; m_q = 30$. (e) $m_s = 60; m_q = 40$. (f) $m_s = 50; m_q = 50$.

Since BLG-PIAA is a gradient optimization-based method, we further leverage a CNN visualization code[2] to show the gradients change in the pixel level of input images in PIAA tasks. We randomly select some testing images from the FLICKR-AES database and use the training data of users who have rated the images to fine-tune the prior model trained on the training set of FLICKR-AES. Fig. 7 shows the visualized gradients at the pixel level (gradient map) computed by the prior model and three users' PIAA models. The aesthetic ratings of users are shown below each gradient map. As can be seen, the gradients of the user's PIAA model are more concentrated in salient regions than that of the prior model. Furthermore, users with different aesthetic ratings on an image have different areas of interest. This demonstrates that the aesthetic prior model can be easily adapted to different users' unique aesthetic preferences on images. In order to verify the fast adaptability of the learned prior model, we show the gradient maps of the image (the first row of Fig. 7) during the fine-tuning of a user's PIAA task on the prior model in Fig. 8. As can be seen, our proposed model can capture the user's area of interest quite well after only five epochs fine-tuning. We have done extensive experiments on diversified images for different users, and have obtained very similar results. This demonstrates that the learned prior model has the ability of fast adaptation for the PIAA task.

In order to check whether the proposed model is subject to overfitting, we conduct experiments on PIAA tasks for all 37 testing users of the FLICKR-AES database and find that the training loss and testing loss can converge well during fine-tuning of their PIAA tasks on the prior model. For example, we show the curves of loss during training and testing versus the number of epochs in fine-tuning for two different users' PIAA models in Fig. 9. For both users, the loss of training and testing can be quickly reduced to a stable value after about 15 epochs fine-tuning on the aesthetic prior model. We conclude that our aesthetic prior model has a strong generalization ability for the PIAA task, which in turn verifies that our model does not overfit.

To further verify that the ability of our prior model in capturing the shared rule that people judge image aesthetics, we conduct a comparative experiment between the PIAA model fine-tuned from the proposed prior model and the PIAA model

[2]https://github.com/sar-gupta/convisualize_nb

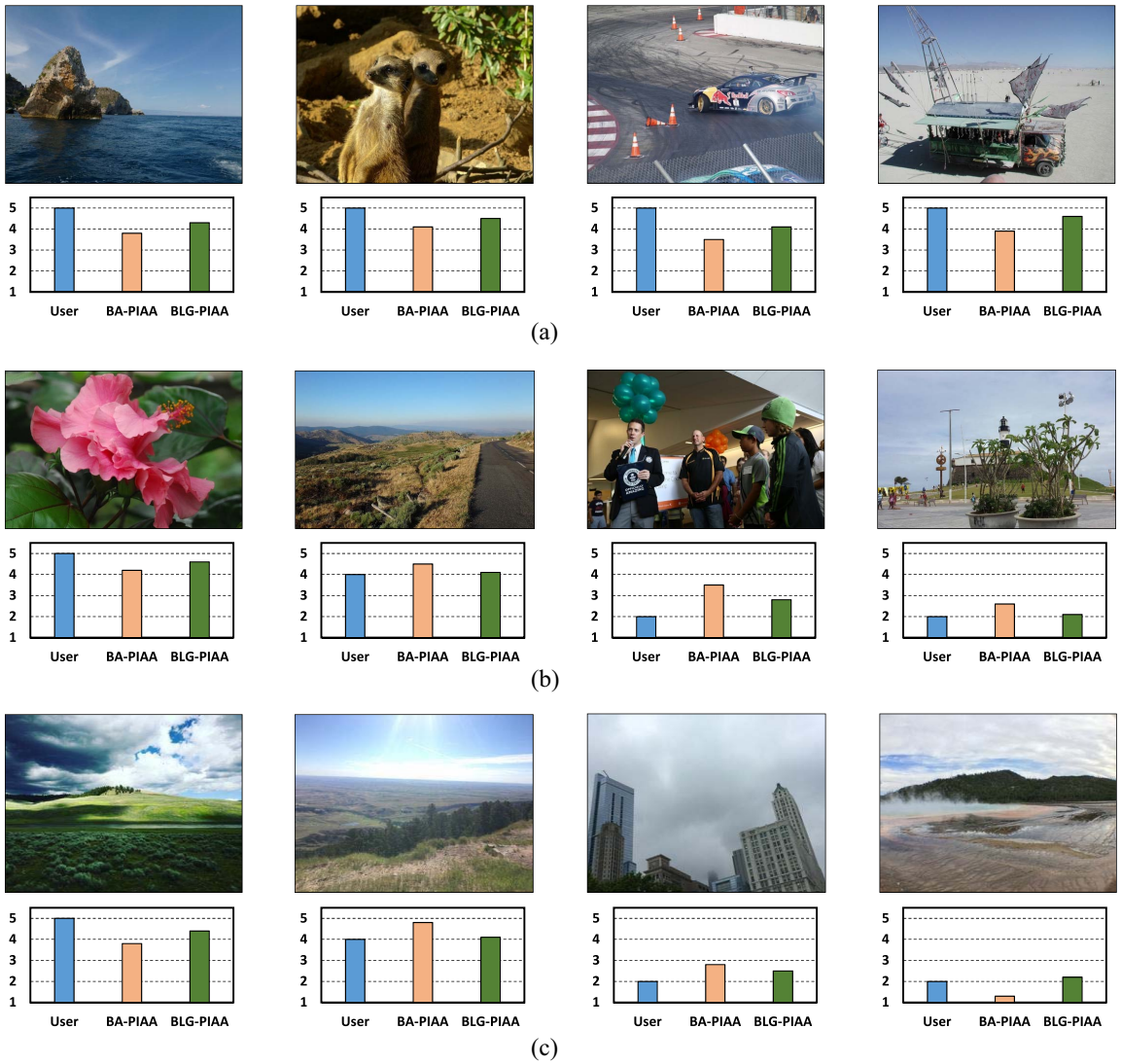


Fig. 6. Some example images rated by three users from FLICKR-AES, AADB, and REAL-CUR. The aesthetic scores rated by the user and the predicted personalized aesthetic scores of BA-PIAA and BLG-PIAA are shown below each image. Example images rated by a user from (a) FLICKR-AES, (b) AADB, and (c) REAL-CUR.

learned directly from users' respective training data. Fig. 10 shows the gradient maps of the image (input image in the second row of Fig. 7) rated by three different users, which are generated based on the above two PIAA models (first column: prior model; second column: respective model). The PIAA model (prior) can be obtained by only 20 epochs fine-tuning on the proposed prior model, while the PIAA model (respective) needs at least 50 epochs training. For any of the three users, the gradient maps of the testing image computed by these two PIAA models are quite similar. This demonstrates that our prior model can effectively adapt to different users' unique aesthetic preferences on images, which in turn demonstrate that our prior model can capture the shared aesthetic knowledge of different users.

As observed from Fig. 4, for some users, the proposed model can only achieve moderate prediction performances. To analyze the reason why the proposed method cannot work well for these users, we select one of the testing users from the FLICKR-AES database for analysis and show the predicted results of four example images rated by the user. Fig. 11 shows the user's ground-truth scores and the predicted scores of BA-PIAA and BLG-PIAA for these images. From Fig. 11, we notice that the user rated the same aesthetic score, that is, 3, to different images. This makes it difficult for our model to learn the user's unique aesthetic preferences on images. In addition, our PIAA model is fine-tuned on the meta-learning-induced prior model with shared aesthetic knowledge. If a user's visual aesthetic preferences are significantly different from the shared aesthetics of the majority of people, the PIAA model cannot accurately infer the user's aesthetic perception of images. Therefore, proposed method relies on users' unique-annotated aesthetic data. If a user cannot provide accurate and diverse-annotated aesthetic ratings on images, the prediction performance of the user can be compromised.

### G. Limitations

While the proposed method has achieved the best performance when compared to the state of the arts, there are still some potential limitations. The proposed method still

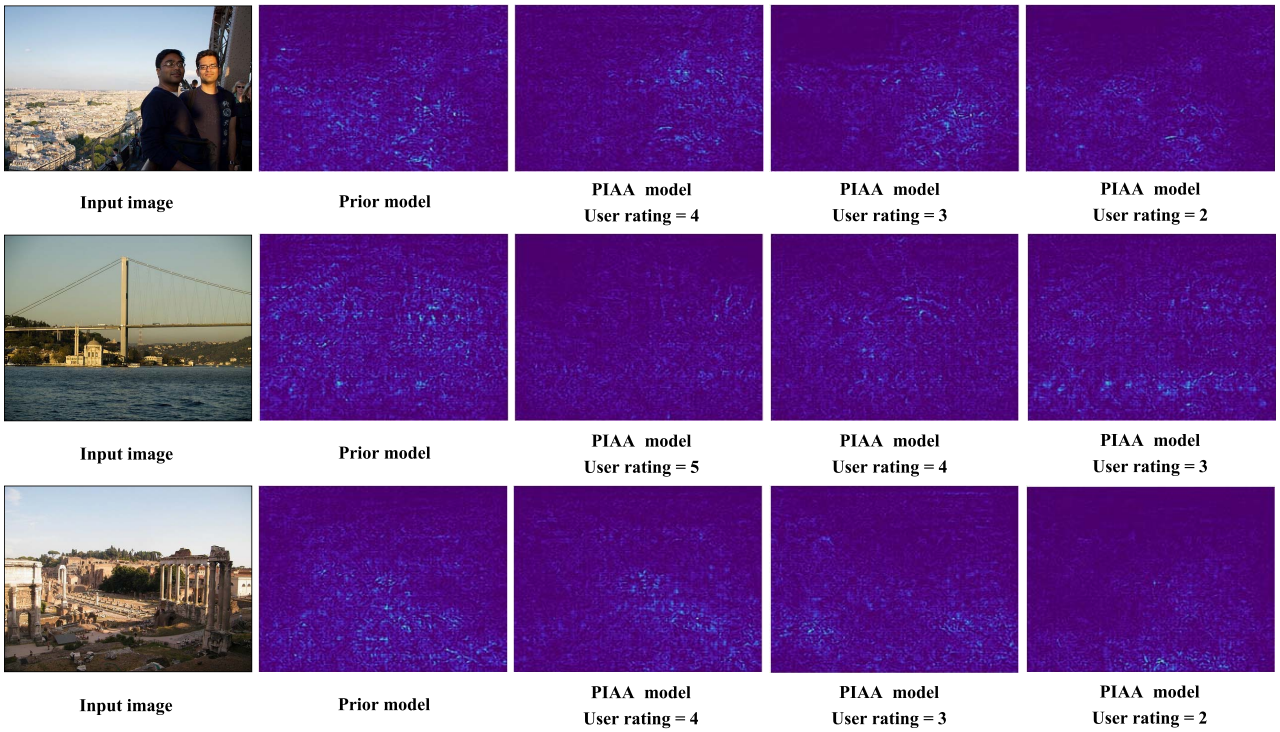


Fig. 7. Some testing images rated by different users from FLICKR-AES. The visualized gradients at the pixel level (gradient map) computed by the prior model and three users' PIAA models are shown beside each image.

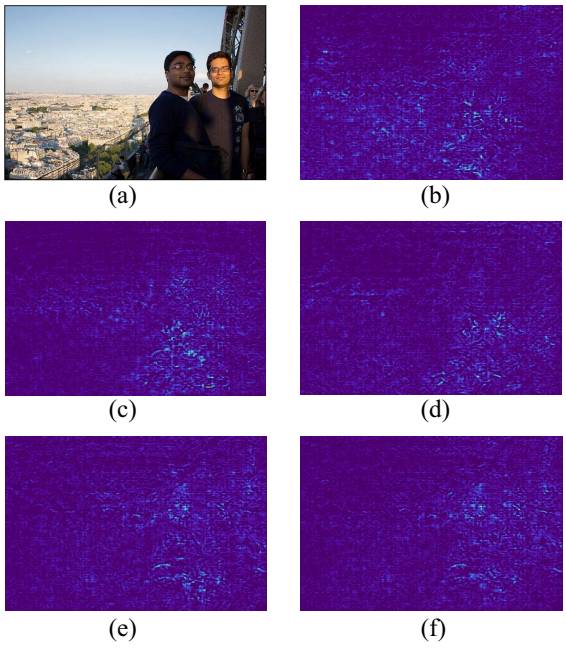


Fig. 8. Testing image in the first row of Fig. 7 and its gradient maps during the fine-tuning of a user's PIAA task on the prior model. (a) Input image. (b) Prior model. (c) After 5 epochs fine-tuning. (d) After 10 epochs fine-tuning. (e) After 15 epochs fine-tuning. (f) After 20 epochs fine-tuning.

needs user-annotated images for model training, and the number of annotated images is relatively large, which in turn brings difficulty in some related practical applications, such as personalized recommendation systems. Therefore, an ideal PIAA method needs to reduce the number of user-annotated images, or even work without using user-annotated data, which can be

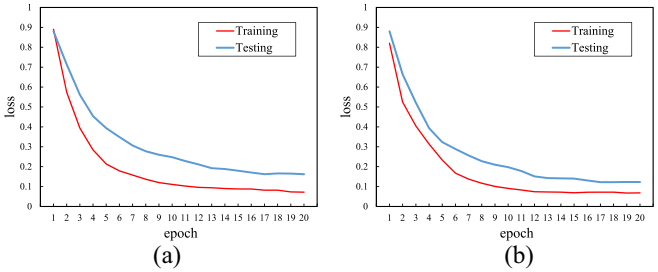


Fig. 9. Loss in training and testing versus the number of epochs in fine-tuning for two different users' PIAA models. (a) User 1. (b) User 2.

achieved by making full use of image date related to users from social networks. In this way, users' unique aesthetic preferences on images can be inferred without the need for aesthetic-annotated data.

Another issue that we have to pay attention to is that, although our PIAA method can achieve the best performance, the prediction accuracy is still moderate (from 0.561 to 0.669). This is mainly because that the PIAA problem is highly subjective and users only provide a small amount of image aesthetic data, which makes PIAA extremely challenging. Furthermore, for individual users, the aesthetic rating of images may be affected by multiple influencing factors (e.g., personality traits [21], [22] and emotions [24], [25]). Therefore, in addition to users' annotated aesthetic data, a better approach should take into account more diversified factors that affect users' aesthetic perception of images. So in the future work, a very promising direction is to incorporate diversified factors related to individual users' image aesthetic perception into PIAA approaches.

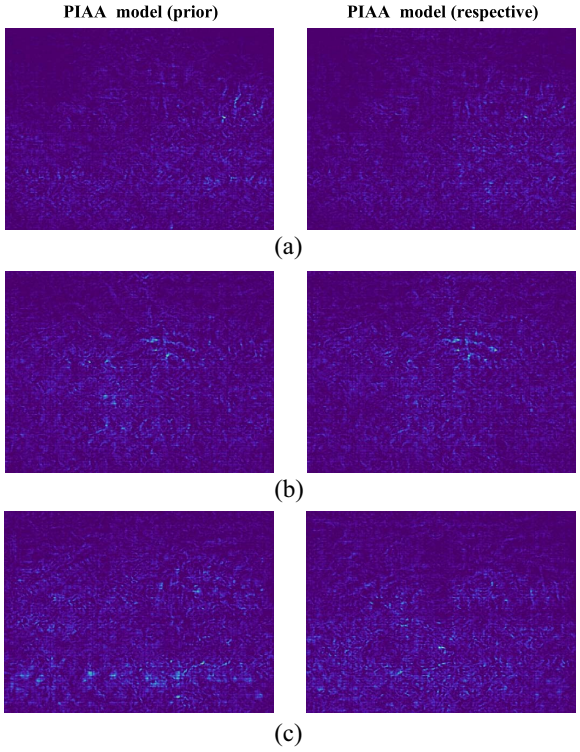


Fig. 10. Gradient maps of the input image in the second row of Fig. 7 computed by two PIAA models of three different users, where the PIAA model (prior) represents the model fine-tuned from the prior model and the PIAA model (respective) denotes the model learned directly from users' respective training data. (a) User 1. (b) User 2. (c) User 3.

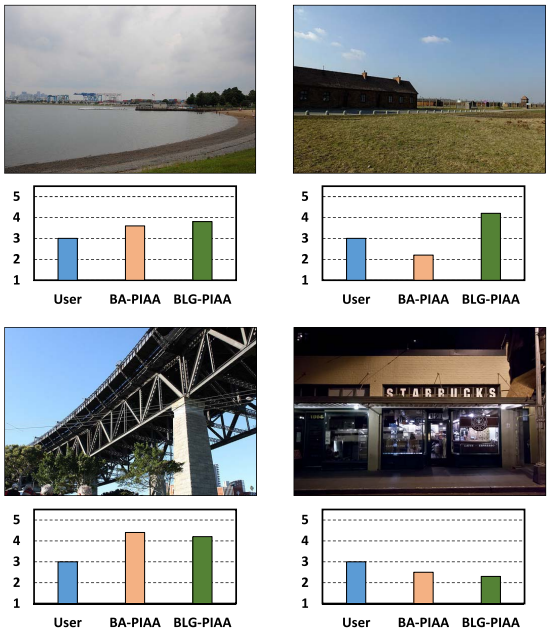


Fig. 11. Four example images rated by a testing user from the FLICKR-AES database. The aesthetic scores rated by the user and the predicted personalized aesthetic scores of BA-PIAA and BLG-PIAA are shown below each image.

## V. Conclusion

In this article, we have proposed a novel PIAA method based on meta-learning with bilevel gradient optimization (BLG-PIAA). Different from the previous PIAA approaches, we have introduced a meta-learning method to solve the SSL problem of PIAA. In our approach, the training data of extensive PIAA tasks were split into support sets and query sets and a bilevel gradient optimization from the support set to the query set was used to learn an effective aesthetic prior model. It can capture the shared rule that people judge image aesthetics and has strong generalization performance for new PIAA tasks. Experiments conducted on three public databases have corroborated that BLG-PIAA outperforms the state-of-the-art PIAA methods. Compared with the baseline approach BA-PIAA based on average aesthetics, BLG-PIAA has proved to be more effective in PIAA tasks without additional network parameters. In addition, experimental results on several basic backbones have demonstrated that BLG-PIAA is an extensible approach that can be applied to the most deep regression networks.

## References

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

[2] T. Li, B. Ni, M. Xu, M. Wang, Q. Gao, and S. Yan, "Data-driven affective filtering for images and videos," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2336–2349, Oct. 2015.

[3] W.-T. Sun, T.-H. Chao, Y.-H. 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.

[4] 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.

[5] R. Hong, L. Zhang, and D. Tao, "Unified photo enhancement by discovering aesthetic communities from Flickr," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1124–1135, Nov. 2016.

[6] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.

[7] Z. Lin, G. Ding, J. Han, and J. Wang, "Cross-view retrieval via probability-based semantics-preserving hashing," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4342–4355, Dec. 2017.

[8] C. Deng, E. Yang, T. Liu, J. Li, W. Liu, and D. Tao, "Unsupervised semantic-preserving adversarial hashing for image search," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4032–4044, Aug. 2019.

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

[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.*, May 2012, pp. 2408–2415.

[11] Y. Chen, Y. Hu, L. Zhang, P. Li, and C. Zhang, "Engineering deep representations for modeling aesthetic perception," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 3092–3104, Nov. 2018.

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

[13] L. Li, H. Zhu, S. Zhao, G. Ding, H. Jiang, and A. Tan, "Personality driven multi-task learning for image aesthetic assessment," in *Proc. IEEE Conf. Multimedia Expo*, Jul. 2019, pp. 430–435.

[14] X. Jin *et al.*, "Predicting aesthetic score distribution through cumulative Jensen–Shannon divergence," in *Proc. 32th AAAI Int. Conf. Artif. Intell.*, Oct. 2018, pp. 77–84.

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

[16] 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.

[17] 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.

[18] Y. Zhu, S. C. Guntuku, W. Lin, G. Ghinea, and J. A. Redi, "Measuring individual video QoE: A survey, and proposal for future directions using social media," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 2s, pp. 1–24, 2018.

[19] E. V. Loosbroek and A. W. Smitsman, "Visual perception of numerosity in infancy," *Develop. Psychol.*, vol. 26, no. 6, pp. 916–922, 1990.

[20] W. Kim, J. Choi, and J. Lee, "Objectivity and subjectivity in aesthetic quality assessment of digital photographs," *IEEE Trans. Affect. Comput.*, early access, doi: 10.1109/TAFFC.2018.2809752.

[21] 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.

[22] 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.

[23] 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.*, to be published, doi: 10.1007/s11063-019-09987-7.

[24] S. Zhao, H. Yao, Y. Gao, G. Ding, and T.-S. Chua, "Predicting personalized image emotion perceptions in social networks," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 526–540, Oct.–Dec. 2018.

[25] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "EmotionMeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019.

[26] P. Lv *et al.*, "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] 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.

[29] 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.

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

[31] J. Shu, Z. Xu, and D. Meng, "Small sample learning in big data era," Aug. 2018. [Online]. Available: http://arxiv.org/abs/1808.04572.

[32] J. Vanschoren, "Meta-learning: A survey," Oct. 2018. [Online]. Available: http://arxiv.org/abs/1810.03548.

[33] Y. Guo, G. Ding, J. Han, and Y. Gao, "Zero-shot learning with transferred samples," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3277–3290, Jul. 2017.

[34] Y. Wang and Q. Yao, "Few-shot learning: A survey," Apr. 2019. [Online]. Available: http://arxiv.org/abs/1904.05046.

[35] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, Aug. 2017, pp. 1126–1135.

[36] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 539–546.

[37] 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.

[38] C. Deng, Y. Xue, X. Liu, C. Li, and D. Tao, "Active transfer learning network: A unified deep joint spectral–spatial feature learning model for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1741–1754, Mar. 2019.

[39] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[40] Y. Zhou, S. Huo, W. Xiang, C. Hou, and S. Kung, "Semi-supervised salient object detection using a linear feedback control system model," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1173–1185, Apr. 2019.

[41] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. P. Lillicrap, "Meta-learning with memory-augmented neural networks," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2016, pp. 1842–1850.

[42] T. Munkhdalai and H. Yu, "Meta networks," in *Proc. Int. Conf. Mach. Learn.*, Aug. 2017, pp. 2554–2563.

[43] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 4080–4090.

[44] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.

[45] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," Oct. 2018. [Online]. Available: http://arxiv.org/abs/1803.02999.

[46] L. Franceschi, P. Frasconi, S. Salzo, R. Grazzi, and M. Pontil, "Bilevel programming for hyperparameter optimization and meta-learning," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2018, pp. 1563–1572.

[47] B. Colson, P. Marcotte, and G. Savard, "An overview of bilevel optimization," *Ann. Oper. Res.*, vol. 153, no. 1, pp. 235–256, Apr. 2007.

[48] Y. Lee and S. Choi, "Gradient-based meta-learning with learned layer-wise metric and subspace," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2018, pp. 2933–2942.

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

[50] S. Ruder, "An overview of gradient descent optimization algorithms," Sep. 2016. [Online]. Available: http://arxiv.org/abs/1609.04747.

[51] 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.

[52] 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.

[53] A. Paszke *et al.*, "Automatic differentiation in Pytorch," in *Proc. Adv. Neural Inf. Process. Syst. Workshop*, Dec. 2017, pp. 1–4.

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

[55] J. L. Myers, A. D. Well, and R. F. Lorch, *Research Design and Statistical Analysis*. New York, NY, USA: Routledge, 2013.


**Hancheng Zhu** received the B.S. degree from the Changzhou Institute of Technology, Changzhou, China, in 2012, and the M.S. degree from the China University of Mining and Technology, Xuzhou, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Information and Control Engineering.

His research interests include affective computing and image aesthetics assessment.

**Leida Li** 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, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search Laboratory, 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 was with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China, as an Assistant Professor, an Associate Professor, and a Professor, respectively. He is currently a Professor with the School of Artificial Intelligence, Xidian University. His research interests include multimedia quality assessment, affective computing, information hiding, and image forensics.

Prof. Li has served as an SPC for IJCAI 2019–2020, the Session Chair for ICMR in 2019 and PCM in 2015, and the TPC for AAAI in 2019, ACM MM 2019–2020, ACM MM-Asia in 2019, ACII in 2019, and PCM in 2016. He is currently an Associate Editor of the *Journal of Visual Communication and Image Representation* and the *EURASIP Journal on Image and Video Processing*.

**Jinjian Wu** received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively.

From 2011 to 2013, he was a Research Assistant with Nanyang Technological University, Singapore, where he was a Postdoctoral Research Fellow from 2013 to 2014. From 2015 to 2019, he was an Associate Professor with Xidian University, where he has been a Professor since 2019. His research interests include visual perceptual modeling, biomimetic imaging, quality evaluation, and object detection.

Prof. Wu received the Best Student Paper Award at ISCAS 2013. He has served as an Associate Editor for the *Journal of Circuits, Systems and Signal Processing*, the Special Section Chair for IEEE Visual Communications and Image Processing in 2017, and the Section Chair/Organizer/TPC Member for ICME 2014–2015, PCM 2015–2016, ICIP 2015, VCIP 2018, and AAAI 2019.

**Sicheng Zhao** received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2016.

He was a Visiting Scholar with the National University of Singapore, Singapore, from 2013 to 2014, and a Research Fellow with Tsinghua University, Beijing, China, from 2016 to 2017. He is currently a Research Fellow with the University of California at Berkeley, Berkeley, CA, USA. His research interests include affective computing, multimedia, and computer vision.

**Guiguang Ding** received the Ph.D. degree in electronic engineering from Xidian University, Xi'an, China, in 2004.

In 2006, he joined the School of Software, Tsinghua University, Beijing, China, where he was a Postdoctoral Research Fellow with the Department of Automation. He is currently an Associate Professor with the School of Software, Tsinghua University. He has published 80 papers in major journals and conferences, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, SIG IR, AAAI, ICML, IJCAI, CVPR, and ICCV. His current research centers on the areas of multimedia information retrieval, computer vision, and machine learning.

**Guangming Shi** received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively.

He had studied with the University of Illinois at Champaign, Champaign, IL, USA, and with the University of Hong Kong, Hong Kong. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University, where he is currently the Academic Leader on circuits and systems. He has authored and coauthored more than 200 papers in journals and conferences. His research interests include compressed sensing, brain cognition theory, multirate filter banks, image denoising, low-bitrate image and video coding, and the implementation of algorithms for intelligent signal processing.

Prof. Shi was awarded the Cheung Kong Scholar Chair Professor by the Ministry of Education in 2012. He served as the Chair for the 90th MPEG and the 50th JPEG of the international standards organization, and the Technical Program Chair for FSKD06, VSPC in 2009, the IEEE Pulse Code Modulation in 2009, the SPIE Visual Communications and Image Processing in 2010, and the IEEE International Symposium on Circuits and Systems in 2013.