Transductive Aesthetic Preference Propagation for Personalized Image Aesthetics Assessment

Yaohui Li[*]
Department of Control Science and Intelligence Engineering
Nanjing University
yaohuili@smail.nju.edu.cn

Yuzhe Yang
OPPO Research Institute
ippllewis@gmail.com

Huaxiong Li[†]
Department of Control Science and Intelligence Engineering
Nanjing University
huaxiongli@nju.edu.cn

Haoxing Chen
Department of Control Science and Intelligence Engineering
Nanjing University
haoxingchen@smail.nju.edu.cn

Liwu Xu
OPPO Research Institute
xuliwu@oppo.com

Leida Li
School of Artificial Intelligence
Xidian University
ldli@xidian.edu.cn

Yaqian Li[†]
OPPO Research Institute
liyaqian@oppo.com

Yandong Guo
OPPO Research Institute
yandong.guo@live.com

ABSTRACT

Personalized image aesthetics assessment (PIAA) aims at capturing individual aesthetic preference. Fine-tuning on personalized data has been proven to be effective in PIAA task. However, a fixed fine-tuning strategy may cause under/over-fitting on limited personal data and it also brings additional training cost. To alleviate these issues, we employ a meta learning-based Transductive Aesthetic Preference Propagation (TAPP-PIAA) algorithm under regression manner to substitute the fine-tuning strategy. Specifically, each user's data is regarded as a meta-task and spilt into support and query set. Then, we extract deep aesthetic features with a pre-trained generic image aesthetic assessment (GIAA) model. Next, we treat image features as graph nodes and their similarities as edge weights to construct an undirected nearest neighbor graph for inference. Instead of fine-tuning on support set, TAPP-PIAA propagates aesthetic preference from support to query set with a predefined propagation formula. Finally, to learn a generalizable aesthetic representation for various users, we optimize our TAPP-PIAA across different users with meta-learning framework. Experimental results indicate that our TAPP-PIAA can surpass the state-of-the-art methods on benchmark databases.

CCS CONCEPTS

• **Computing methodologies → Image representations**.

KEYWORDS

personalized image aesthetic assessment; meta-learning; transductive label propagation

ACM Reference Format:
Yaohui Li, Yuzhe Yang, Huaxiong Li, Haoxing Chen, Liwu Xu, Leida Li, Yaqian Li, and Yandong Guo. 2022. Transductive Aesthetic Preference Propagation for Personalized Image Aesthetics Assessment. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), Oct. 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503161.3548244

1 INTRODUCTION

Computational photo aesthetic assessment is a long-standing problem in affective computing research [24]. With explosive growth of camera phones and self-media market, the ability of photo management has shown great essentiality in many real-world problems [6]. However, picking up higher quality and appealing photos manually is time-consuming and tangled usually. To save time and be more efficient, we hope that our computational machines could share similar aesthetic perception as human and filter materials automatically. Traditional image aesthetic assessment aims at generating predictions based on voting from multi-annotators or mean opinion scores (MOS) [15], it is also known as generic image aesthetic assessment (GIAA). The GIAA results usually are not close to each individual preference, since it represents "average opinion". Hence, it works well in low-quality material filtering, photo pre-processing and etc. However, as is mentioned above, each person has an individual "aesthetic taste", which is usually not in line with the average result [36]. If we directly apply the GIAA results to personalized photo recommendation, it usually brings limited user experience. To this end, personalized image aesthetic assessment (PIAA) is proposed for capturing aesthetic preferences among different individuals [25].

However, learning to model unique aesthetic tastes for different individuals is cumbersome. 1) First, human aesthetic perception on

[*] This work was done during the internship at OPPO Research Institute.
[†] Corresponding authors.

MM '22, October 10–14, 2022, Lisboa, Portugal
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3548244

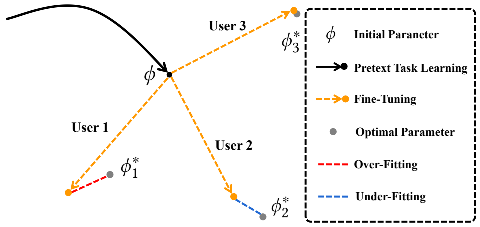

Figure 1: Learning individual aesthetic preferences by fine-tuning on limited data may cause under/over-fitting. In this work, we utilize a graph-based transductive knowledge propagation algorithm instead of fine-tuning.

single image varies from one to another. Therefore, it requires the learning model to have a strong generalization ability, so that it can fit well on various users' aesthetic preferences. 2) Second, the amount of annotated personalized data is often limited [25] and learning with scarce data is a fundamental challenge in deep model-based vision reasoning. To summarize, PIAA requires the model to have a strong generalization ability while essentially we only have limited annotated data. The aforementioned requirements are in line with few-shot learning (FSL) settings, which inspire us to locate PIAA under the FSL scenario.

Few-shot learning aims at designing efficient learning algorithms that can generalize to novel tasks with limited annotations [7]. Recently, meta learning-based approaches have achieved success in FSL. Specifically, meta-learning utilizes the episodic training mechanism to obtain a representation with strong generalization ability [28]. By training on a series of meta-tasks, meta-learning has been proven to generalize well in many vision tasks, such as few-shot image classification [3, 17, 34]. Beyond these, recent works [31, 36] have also introduced meta-learning into PIAA task. By regarding each user's data as a meta-task, learning algorithms are designed to capture individual aesthetic preference by training across different users.

Although meta-learning framework can improve the generalization ability across different users, how to learn individual aesthetic preference sufficiently with scarce data remains challenging. Existing methods learn user preferences with limited data by a fixed fine-tuning strategy, which may cause under/over-fitting [1, 2] (see Fig. 1) and also bring additional training cost. To address these issues, we introduce a meta learning-based Transductive Aesthetic Preference Propagation (TAPP-PIAA) algorithm into PIAA task. Instead of directly training with limited annotations, we propose to make use of unlabeled data with transductive inference and propagate individual aesthetic knowledge through an undirected graph. First, we split each meta-task into a labeled support set and an unlabeled query set. Then, we extract all image features with a GIAA pre-trained CNN. To model relationship among images, we regard image features as nodes and their similarities as edge weights to construct an undirected nearest neighbor graph. Based on transductive inference, we propagate individual aesthetic knowledge among connected graph nodes with a predefined propagation formula. Finally, we optimize our TAPP-PIAA algorithm under the meta-learning framework to learn a representation with stronger generalization abilities.

Contributions of this work are summarized in three folds:

- We first-time propose using a Transductive Aesthetic Preference Propagation (TAPP-PIAA) algorithm that introduces the graph-based label propagation mechanism into PIAA task. It takes advantage of both labeled and unlabeled data to infer aesthetic preferences without personalized fine-tuning.
- We extend the existing classification-based label propagation mechanism into a regression-based one for modeling aesthetic preferences in a continuous space. The extended propagation mechanism is also applicable in meta-learning framework to enhance the generalization ability.
- We apply the proposed TAPP-PIAA on three benchmark databases Flickr-AES, AADB and REAL-CUR. Experimental results show that our TAPP-PIAA can generalize well on unseen users and outperform the state-of-the-art method.

2 RELATED WORKS

2.1 Personalized Image Aesthetic Assessment

The approaches of modeling aesthetic preference varies. [25] proposes a residual-based method to model PIAA task by aggregating generic and personal preference offset together. Multi-modal learning has also been introduced into PIAA task in [30], which takes advantage of text-based photo aesthetic reviews to obtain personalized aesthetic representations. Beyond these, inspired by multi-task learning strategy, recent work PA-IAA [15] proposes jointly learning personalized and generic aesthetic data to boost the performances of both GIAA and PIAA. Meanwhile, [19] has adapted deep reinforcement learning strategies into PIAA task, which makes use of users' interactions for training.

In addition, meta-learning framework has also been introduced into PIAA task for it can enhance the generalization ability of models [8, 9]. [31] first-time introduces meta-learning framework into PIAA task and models personalized residual components to assist the pre-trained GIAA model. In this way, model can learn to infer aesthetic preferences of various users and generalize well to novel users. Besides, in order to learn aesthetic prior knowledge with a stronger generalization ability, BLG-PIAA [36] utilizes a bilevel gradient optimization strategy [9] to optimize the model under the meta-learning framework. With the meta-learned prior knowledge, the model can generalize to unseen users quickly. Although meta-learning framework has achieved promising performances in PIAA task, the fundamental difficulty of mining individual aesthetic preference with limited annotations has not yet been properly solved. Most existing methods learn individual preference by fine-tuning [15, 25, 31, 36]. In this work, we follow the meta-learning framework and focus on digging out personal aesthetic preferences with the help of unlabeled data.

2.2 Meta-Learning

Despite the great success in deep learning, large-scale learning algorithms still suffer from their data-driven natures and can hardly generalize to unseen tasks with a few annotations [32]. To tackle these limitations, meta-learning has been introduced to improve generalization ability with designed episodic training mechanism [28, 31]. Specifically, the training set is divided into many episodes (meta-tasks), where each episode contains a small labeled support

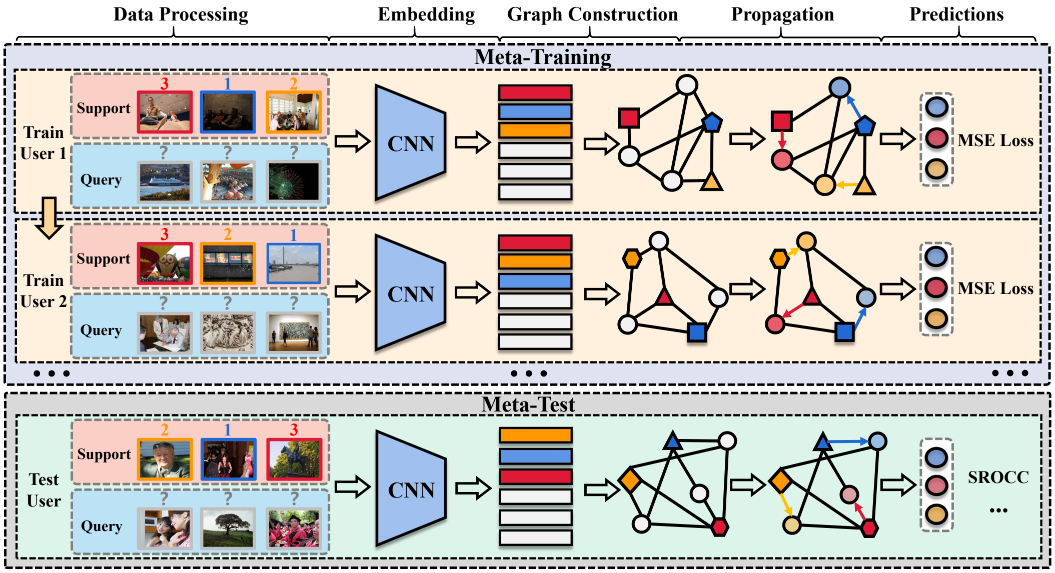

Figure 2: The overall framework of our proposed TAPP-PIAA.

set for training and an unlabeled query set for evaluation. Then, a series of meta-tasks are feed into learning algorithms to learn task adaptation knowledge. Therefore, episodic training mechanism can enhance the generalization ability and help the algorithm to learn novel tasks with few data [8, 16].

Existing meta-learning approaches could be briefly divided into two branches: optimization-based and metric learning-based. First, optimization-based methods aim to learn shared prior knowledge that can generalize to novel tasks quickly [8, 9]. Specifically, MAML [8] utilizes a model-agnostic meta-learner to learn shared aesthetic prior knowledge via a second-order gradient optimization. Besides, [9] regards meta-training process as a hierarchical optimization problem and thus simplifies the optimization procedures. Second, metric learning-based methods aim to learn a shared representation space for different tasks [26]. Different from optimization-based methods, without training on scarce data, metric learning-based methods make inference according to the similarities between support and query samples. Typical methods like DN4 [16] represents images by deep local features and designs an image-to-class measurement strategy to compare pair-wise semantic similarities. Besides, TPN [18] incorporates transductive inference into meta-learning framework to deal with the problem of data scarcity by making use of both labeled and unlabeled data. Based on above discussions and analyses, we propose to explore PIAA task from the perspective of metric learning-based meta-learning. Moreover, to deal with data scarcity, we decide to make use of unlabeled query samples through the transductive inference.

2.3 Transductive Inference

Transductive inference [27] is one of the semi-supervised learning methods. Inductive inference first induces a rule from training set and then uses it to predict test set. Differently, transductive inference follows the assumption that test set is available. As a result, transductive inference can alleviate data scarcity to a certain extent, so it is suitable for few-shot tasks. In this work, we follow the transductive inference to tackle PIAA task. Transductive label propagation, as a main mean of transductive inference, utilizes a graph structure to discover the relationship between labeled and unlabeled samples. Early works [35] and [37] propagate label information from labeled to unlabeled samples via a fully-connected graph. In few-shot scenarios, TPN [18] and iLPC [14] first extract support and query features and then construct an undirected nearest neighbor graph. Then, classification knowledge is propagated from support to query samples through graph edges. In this work, we modify the traditional classification-based transductive label propagation to a regression-based one for aesthetic preference propagation due to its subjective nature.

3 PROPOSED METHOD

3.1 Problem Formulation

In our proposed method, we follow the meta-learning framework [28] to divide the whole pipeline into two parts: meta-training $\mathcal{T}_{tr}$ and meta-test $\mathcal{T}_{te}$. In meta-training phase, we sample T_r meta-tasks from training users’ data: $\mathcal{T}_{tr} = \{(\mathcal{D}^s_{tr}, \mathcal{D}^q_{tr})_i\}^{T_r}_{i=1}$. Here, $(\mathcal{D}^s_{tr}, \mathcal{D}^q_{tr})_i$ represents the pair of training and test sets in the i-th meta-task, which are usually called support and query sets. Specifically, $\mathcal{D}^s_{tr}$ contains N labeled samples while $\mathcal{D}^q_{tr}$ includes M unlabeled samples. Note that $\mathcal{D}^s_{tr}$ and $\mathcal{D}^q_{tr}$ from the same meta-task are both sampled from the same user’s data without intersection. In meta-test phase, we follow the settings of meta-training and sample totally T_e meta-tasks from test users’ data: $\mathcal{T}_{te} = \{(\mathcal{D}^s_{te}, \mathcal{D}^q_{te})_j\}^{T_e}_{j=1}$ for evaluation. Under meta-learning framework, we optimize our model on $\mathcal{T}_{tr}$ to boost its performance on unseen novel meta-tasks in $\mathcal{T}_{te}$. The objective function for meta-training is defined as follow:

$$\min_{\phi} \mathop{\mathbb{E}}_{\substack{\mathcal{T}_{tr}\sim p(\mathcal{T}) \\ (\mathcal{D}^s_{tr},\mathcal{D}^q_{tr})\in\mathcal{T}_{tr}}} \sum_{(x,y)\in\mathcal{D}^q_{tr}} \mathcal{L}(\zeta(\mathcal{F}_{\phi}(x), \mathcal{F}_{\phi}(\mathcal{D}^s_{tr})), y; \phi), \qquad (1)$$

where $p(\mathcal{T})$ is the overall task distribution. Besides, $\mathcal{F}_\phi$ is an embedding network with learnable parameter ϕ and $\zeta(\cdot)$ is a reasoning strategy to predict the label of x. The prediction is then used to calculate loss function $\mathcal{L}$ with ground truth y. In this work, $\zeta(\cdot)$ is our TAPP-PIAA algorithm.

3.2 Overall Framework

As is shown in Fig. 2, we divide the whole pipeline into two phases: meta-training and meta-test. Notice that they both consist of randomly sampled meta-tasks. Meta-training phase contains four parts: feature embedding, graph construction, transductive aesthetic preference propagation and loss computation. Specifically, given a meta-task, we first extract features of all samples. Then, we regard all image features in a meta-task as nodes and their similarities as edge weights to construct an undirected nearest neighbor graph. After that, aesthetic preference is directly propagated with a designed rule. Finally, the mean square error loss is utilized for optimization. During meta-training phase, our model is trained across a series of meta-tasks to learn a generalizable aesthetic representation. During meta-test phase, we utilize the meta-trained model to infer test user's aesthetic preference.

3.3 Feature Embedding

Given an input image x, we utilize a pre-trained CNN $\mathcal{F}_\phi$ with learnable parameter ϕ to extract its feature map as a $3D$ tensor:

$$\hat{x} = \mathcal{F}_\phi(x) \in \mathbb{R}^{c \times h \times w}, \tag{2}$$

where c, h and w are the three dimensions of $\hat{x}$. Given a meta-task $(\mathcal{D}^s_{tr}, \mathcal{D}^q_{tr})$, where support set $\mathcal{D}^s_{tr}$ consists of N labeled images $\{x^s_1, \ldots, x^s_N\}$ with labels $\{y^s_1, \ldots, y^s_N\}$ and query set $\mathcal{D}^q_{tr}$ consists of M unlabeled images $\{x^q_1, \ldots, x^q_M\}$. We embed all images by $\mathcal{F}_\phi$ and construct a feature set: $\hat{X} = \{\hat{x}^s_1, \ldots, \hat{x}^s_N, \hat{x}^q_1, \ldots, \hat{x}^q_M\}$.

3.4 Nearest Neighbor Graph Construction

After feature extraction, we detail the procedures of graph construction. Generally, a graph $G = (V, E)$ consists of two parts: a set of vertices V (i.e., nodes) and a set of edges E. Specifically, we first normalize the feature set $\hat{X}$ with l_2-norm to form the vertex set $V = \{v_1, \ldots, v_N, v_{N+1}, \ldots, v_{N+M}\}$. Since we want aesthetic knowledge to propagate among aesthetically similar images [11], we design to construct an undirected nearest neighbor graph [14] for aesthetic preference propagation. Here, a sparse *affinity matrix* $W \in \mathbb{R}^{(N+M)\times(N+M)}$ is defined below based on Cosine similarity:

$$W_{i,j} = \begin{cases} exp(\frac{v_i^{\mathsf{T}} v_j}{2\sigma^2}), & if\, i \neq j \wedge v_i \in \mathcal{N}_k(v_j) \\ 0, & otherwise \end{cases}, \tag{3}$$

$i, j \in [1, N+M]$, where v_i is the i-th element of the l_2-normalized vertex set V and $\mathcal{N}_k(v_j)$ denotes the set of top k nearest neighbors of v_j in V. Besides, σ is a predefined scale parameter of the Gaussian function. With Eq. 3, we can obtain a symmetric non-negative *adjacency matrix* $A = \frac{1}{2}(W + W^{\mathsf{T}})$. Finally, we normalize each row of A and denote the normalized *adjacency matrix* as $\bar{A}$.

3.5 Aesthetic Preference Propagation

According to the high correlation between aesthetic features, traditional classification-based label propagation may not be suitable for PIAA task (quantitatively discussed in Subsection 6.1). Therefore, in this work, we propose to propagate aesthetic preference under regression manner by optimizing a mean square error loss. Specifically, we define the initial label matrix $Y \in \mathbb{R}^{(N+M)\times 1}$ as:

$$Y_i = \begin{cases} y^s_i, & 1 \leq i \leq N \\ 0, & N+1 \leq i \leq N+M \end{cases}, \tag{4}$$

where y^s_i is the ground-truth label of the i-th support image. With the constructed nearest neighbor graph and transductive inference, each node interacts with its neighbors and updates its own label [14]. This means the labels of all nodes change iteratively until convergence, including both support and query labels. For persistent source knowledge from support set, we decide to maintain the initial labels of support set. Here, we clamp the original support labels after each propagation iteration [37].

Specifically, we divide the whole label matrix Y into two parts: $Y = \begin{pmatrix} Y_l \\ Y_u \end{pmatrix}$, where $Y_l \in \mathbb{R}^{N\times 1}$ refers to support labels and $Y_u \in \mathbb{R}^{M\times 1}$ is a zero matrix, denoting the initial labels of query set. Similarly, we divide the normalized *adjacency matrix* $\bar{A}$ into four sub-matrices:

$$\bar{A} = \begin{bmatrix} \bar{A}_{ll} & \bar{A}_{lu} \\ \bar{A}_{ul} & \bar{A}_{uu} \end{bmatrix}, \tag{5}$$

where $\bar{A}_{ll} \in \mathbb{R}^{N\times N}, \bar{A}_{lu} \in \mathbb{R}^{N\times M}, \bar{A}_{ul} \in \mathbb{R}^{M\times N}$ and $\bar{A}_{uu} \in \mathbb{R}^{M\times M}$. Then, we denote the label matrix at t-th step as $f^t = \begin{pmatrix} f^t_l \\ f^t_u \end{pmatrix}$ and define the iterative propagation formula as: $f^{t+1} = \bar{A} f^t, f^0 = Y$, which is further expanded to:

$$\begin{pmatrix} f^{t+1}_l \\ f^{t+1}_u \end{pmatrix} = \begin{bmatrix} \bar{A}_{ll} & \bar{A}_{lu} \\ \bar{A}_{ul} & \bar{A}_{uu} \end{bmatrix} \begin{pmatrix} f^t_l \\ f^t_u \end{pmatrix}, \begin{pmatrix} f^0_l \\ f^0_u \end{pmatrix} = \begin{pmatrix} Y_l \\ Y_u \end{pmatrix}. \tag{6}$$

Based on the above analysis, we clamp Y_l iteratively, which means $f^t_l = Y_l$ holds for all t values. Therefore, the propagation formula can be rewritten as:

$$f^{t+1}_u = \bar{A}_{ul} f^t_l + \bar{A}_{uu} f^t_u = \bar{A}_{ul} Y_l + \bar{A}_{uu} f^t_u. \tag{7}$$

The above formula can lead to a closed-form solution [37]:

$$f^*_u = (I - \bar{A}_{uu})^{-1} \bar{A}_{ul} Y_l, \tag{8}$$

where I is an identity matrix and we can directly use Eq. 8 to calculate the final propagation results in practice.

3.6 Objective Function

After above procedures, we can predict a user's aesthetic preference f^*_u on unlabeled query samples. Finally, we adopt the mean square error loss to optimize each meta-task:

$$L = \frac{1}{2M} ||f^*_u - \bar{Y}_u||^2, \tag{9}$$

where $\bar{Y}_u$ represents the ground-truth labels of query set and M is the number of query samples. In meta-training phase, our model is trained on various meta-tasks to enhance its generalization ability. During meta-test phase, we directly apply the meta-trained model on novel users without additional training. The detailed flow of our TAPP-PIAA is shown in Algorithm 1.

Algorithm 1 Algorithm Flow of TAPP-PIAA

```
Input: Meta-training set $\mathcal{T}_{tr}$; meta-test set $\mathcal{T}_{te}$; learning rate $\beta$;
       embedding network $\mathcal{F}_{\phi}$
Output: Predicted labels $f_u^*$ of query set in each meta-task
 1: Initialize model parameters $\phi$.
    Meta-Training Phase
 2: for each meta-task $(\mathcal{D}_{tr}^s, \mathcal{D}_{tr}^q)$ in $\mathcal{T}_{tr}$ do
 3:   while no converge do
 4:     $\hat{X} \leftarrow \mathcal{F}_{\phi}(\mathcal{D}_{tr}^s \cup \mathcal{D}_{tr}^q)$;      ▷ feature extraction
 5:     $V \leftarrow \hat{X}$;      ▷ $l_2$-normalization
 6:     for $i, j \in [1, N+M]$ do $W_{ij} \leftarrow$ affinity values;  ▷ by Eq. 3
 7:     $A = \frac{1}{2}(W + W^{\top})$;      ▷ symmetric adjacency matrix
 8:     $\bar{A} = normalize(A)$;      ▷ graph normalization
 9:     $f_u^* \leftarrow$ Eq. 8;      ▷ transductive propagation
10:     $\phi \leftarrow \phi - \beta\nabla L$.      ▷ optimization by Eq. 9
11:   end while
12: end for
    Meta-Test Phase
13: for meta-task $(\mathcal{D}_{te}^s, \mathcal{D}_{te}^q)$ in $\mathcal{T}_{te}$ do
14:   Load the optimal meta-trained parameters $\phi^*$;
15:   $\hat{X} \leftarrow \mathcal{F}_{\phi^*}(\mathcal{D}_{te}^s \cup \mathcal{D}_{te}^q)$;      ▷ feature extraction
16:   $f_u^* \leftarrow$ follow step 5-9.      ▷ inference
17:   return $f_u^*$.
18: end for
```

4 EXPERIMENTS

4.1 Databases

In this work, we evaluate the effectiveness of our proposed method and conduct further analyses on three frequently used databases.

FLICKR-AES [25] database consists of 210 users and about 40, 000 images. In this work, 173 users make up the training set and the rest 37 users make up the test set for PIAA task. The number of images annotated by test users ranges from 105 to 171.

AADB [13] database contains almost 10, 000 images rated by 190 users and each image is rated by 5 users. Except aesthetic scores, 11 aesthetic attributes are also provided for each image, which have been proved to be useful when analyzing different aesthetic preferences. Similarly, we spare 68 users with as training users and the rest 22 users for test. The number of images rated by test users ranges from 110 to 190.

REAL-CUR [25] database is a relatively smaller database with 14 users' albums and their rated images. The number of images rated by the test users is around 200. Since REAL-CUR is constructed by users' real photo albums, it can simulate realistic application settings.

4.2 Mythology Details

The proposed TAPP-PIAA consists of four steps: feature extraction, graph construction, transductive aesthetic preference propagation and loss computation. First of all, following [25, 36], we initialize our backbones by parameters pre-trained on ImageNet [5]. In this work, we adopt the meta-learning framework to train our model, which mainly focuses on learning task adaptation knowledge. As a result, the learning algorithm may fail to learn fundamental aesthetic knowledge. To tackle this limitation, we follow [33] and further pre-train the embedding network $\mathcal{F}_{\phi}$ with a supervised regression task on all users' mean opinion scores. Note that the GIAA pre-training has no risk of personalized information leakage for GIAA is invariant with PIAA. Meanwhile, this is also in line with realistic settings where generic aesthetic information is available. After the GIAA pre-training, our model is able to extract generic aesthetic representations effectively.

Then, following the paradigm of meta-learning, our model is meta-trained across many meta-tasks. Since the amount of samples rated by some training users are insufficient, which may cause disturbances during training process. In this work, we choose training users with sufficient samples for meta-training. Specifically, we choose 111 and 68 training users for FLICKR-AES and AADB respectively. Under the meta-training framework with selected training users, our model is forced to learn to propagate personalized aesthetic preferences across various users.

In this work, we adopt the Spearman Rank-Order Correlation Coefficient (SROCC) [21] as our main evaluation criterion. Denote the difference value between the prediction and ground-truth of the i-th query sample as d_i. The SROCC index between M predictions and ground-truth labels is defined as:

$$\text{SROCC} = 1 - \frac{6\sum_{i=1}^{M} d_i^2}{M(M^2-1)}, \tag{10}$$

which ranges from -1 to 1 and higher SROCC represents higher correlation between predictions and ground-truth labels.

4.3 Implementation Details

In this work, we implement our method under PyTorch [23] framework and optimize it with an Adam [12] optimizer. For feature embedding, we use two deep neural networks as backbone structures: ResNet-18 and ResNet-50 [10], respectively. Specially, after the last convolutional block of the backbone, we add an adaptive average pooling layer to squeeze the spatial size of the output features to 1×1. Next, we will state the settings of hyper-parameters used in this work. For the GIAA pre-training, we set the initial learning rate to $1e-4$ and decay it by 0.9 after each epoch before convergence. For the baseline Base-PIAA, we fine-tune the GIAA prior model by replacing its last FC layer on personal data 20 times with learning rate $1e-5$. Then, for meta-training on FLICKR-AES and AADB, we randomly choose 100 training images and 40 test images of each training user to construct a meta-task. Since the backbone is pre-trained, we set the initial learning rate of meta-training to $5e-6$ and decay it by 0.9 after each epoch. Besides, the hyper-parameter σ for graph construction in Eq. 3 is predefined to $\sigma^2 = 0.05$. Moreover, the value of nearest neighbors k in Eq. 3 has a significant influence on model's performance, so we report the results of k=30 for 10-shot and k=50 for 100-shot PIAA task according to the investigation in Subsection 5. Finally, during meta-test phase, we set the number of images in support set to 10 for 10-shot tasks and 100 for 100-shot tasks. Note that the remaining images of each test user are collected in query set. In order to avoid the random errors, we sample on each test user's data 50 times to construct different meta-tasks; for each database, we evaluate our method on all test users 20 times and report the mean results with standard deviation.

Table 1: Experimental results on FLICKR-AES. We report the mean results with standard deviations.

Model	*SROCC*	
	10-shot	***100-shot***
FPMF(only attribute) [22]	0.511±0.004	0.516±0.003
FPMF(only content) [22]	0.512±0.002	0.516±0.010
FPMF(content& attribute) [22]	0.513±0.003	0.524±0.007
PAM(only attribute) [25]	0.518±0.003	0.539±0.013
PAM(only content) [25]	0.515±0.004	0.535±0.017
PAM(content& attribute) [25]	0.520±0.003	0.553±0.012
USAR_PPR [20]	0.521±0.002	0.544±0.007
USAR_PAD [20]	0.520±0.003	0.537±0.003
USAR_PPR& PAD [20]	0.525±0.004	0.552±0.015
PA-IAA [15]	0.543±0.003	0.639±0.011
BA-PIAA [36]	0.524±0.004	0.583±0.014
PIAA(MAML) [36]	0.520±0.005	0.569±0.010
PIAA(Reptile) [36]	0.529±0.006	0.598±0.015
BLG-PIAA [36]	0.561±0.004	0.669±0.013
Base-PIAA (Ours)	0.529±0.006	0.585±0.012
Inductive-PIAA (Ours)	0.560±0.006	0.661±0.013
TAPP-PIAA (Ours)	**0.591±0.007**	**0.685±0.012**

4.4 Experimental Results

In this subsection, we evaluate our proposed method on three benchmark databases: FLICKR-AES, AADB and REAL-CUR.

1) Experimental results on FLICKR-AES: We compare our proposed method with state-of-the-art PIAA methods on FLICKR-AES. As is shown in Table 1, on FLICKR-AES, our TAPP-PIAA outperforms the state-of-the-art method BLG-PIAA [36] by 0.03 and 0.016 under 10-shot and 100-shot tasks, respectively. When compared the baseline model Base-PIAA, our TAPP-PIAA achieves 0.062 and 0.1 improvements on 10-shot and 100-shot PIAA tasks respectively. This indicates the effectiveness of our meta learning-based transductive aesthetic preference propagation algorithm which makes use of both labeled and unlabeled data. Moreover, when compared with the inductive learning method on PIAA task: Inductive-PIAA, our TAPP-PIAA achieves 0.031 and 0.024 improvements under both settings. This further demonstrates the advantage of transductive inference, which first-time makes use of both labeled and unlabeled data of each user in PIAA task.

2) Experimental results on AADB: As is shown in Table 2, our method can also achieve the state-of-the-art performance with 0.037 and 0.067 improvements on AADB database. This further demonstrates the superiority of our transductive aesthetic preference propagation algorithm in PIAA task. Besides, when compared with Inductive-PIAA based on inductive reasoning, our TAPP-PIAA achieves 0.01 and 0.047 improvements on two PIAA tasks, respectively.

3) Experimental results of cross-database evaluations: Under the setting of realistic applications, we expect a PIAA model can effectively generalize to various users. Therefore, we conduct a cross-database evaluation by meta-training and meta-testing on different databases. Specifically, we utilize the TAPP-PIAA model meta-trained on FLICKR-AES and AADB to predict the personalized aesthetic preferences of test users in three databases. We report the results in Table 3. From the results on large-scale databases FLICKR-AES and AADB, we can observe that the two models meta-trained on different databases both achieve better performances when meta-tested on FLICKR-AES. This indicates the test users of FLICKR-AES may give more accurate PIAA annotations, so models can learn their aesthetic preferences more easily. For REAL-CUR, the model meta-trained on FLICKR-AES outperforms the model meta-trained on AADB with 0.038 improvement. This indicates that the model trained on FLICKR-AES has a stronger generalization ability than the one trained on AADB.

Table 2: Experimental results on AADB. We report the mean results with standard deviations.

Model	*SROCC*	
	10-shot	***100-shot***
BA-PIAA [36]	0.450±0.001	0.513±0.005
BLG-PIAA [36]	0.497±0.003	0.545±0.007
Inductive-PIAA (Ours)	0.524±0.003	0.565±0.006
TAPP-PIAA (Ours)	**0.534±0.004**	**0.612±0.007**

Table 3: Experimental results of TAPP-PIAA under the cross-database setting on 100-shot PIAA task.

Training Database	Test Database		
	FLICKR-AES	**AADB**	**REAL-CUR**
FLICKR-AES	0.685	0.540	0.580
AADB	0.615	0.612	0.542

5 ABLATION STUDY

5.1 Backbone Structure

According to previous researches, the effectiveness of deep metric-learning highly depends on the quality of representation learning [4]. Therefore, in this subsection, we propose to investigate the influence of different backbone structures of $\mathcal{F}_{\phi}$. As is shown in Table 4, we choose two widely used ResNet-18 and ResNet-50 [10] for backbone evaluation. We can observe that the performances on 100-shot PIAA task vary with different backbone structures. We can further conclude that a deeper embedding network with a stronger representation ability can lead to a higher performance in PIAA task. Therefore, we report the other experimental results based on ResNet-50 in this work.

5.2 Metric Function

In this subsection, we explore the influence of different metric functions for constructing the undirected nearest neighbor graph. As is introduced in Subsection 3.4, we propose to spread aesthetic preferences through a graph structure. Therefore, the way of graph construction is important to the quality of knowledge propagation,

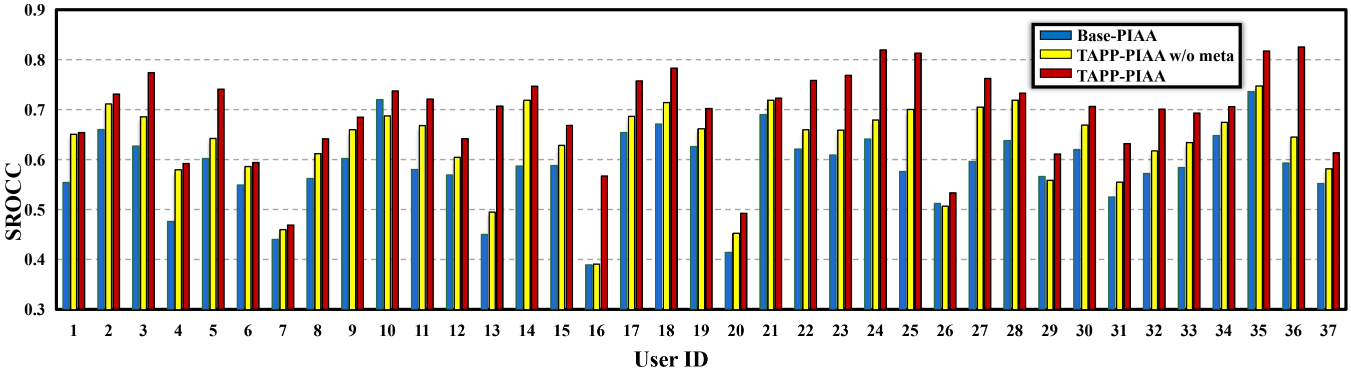

Figure 3: Visualization of 37 test users from FLICKR-AES under the setting of 100-shot PIAA task. The blue, yellow and red bars represent the SROCC values of the baseline Base-PIAA, TAPP-PIAA without meta-training and TAPP-PIAA, respectively.

Table 4: Comparisons of different backbone structures on 100-shot PIAA task.

Backbone	FLICKR-AES		AADB	
	SROCC	*PLCC*	*SROCC*	*PLCC*
ResNet-18	0.641	0.656	0.573	0.591
ResNet-50	**0.685**	**0.704**	**0.612**	**0.627**

Table 5: Comparison of different metric functions for graph construction on 100-shot PIAA task.

Metric	FLICKR-AES		AADB	
	SROCC	*PLCC*	*SROCC*	*PLCC*
Euclidean	0.655	0.662	0.584	0.601
Cosine	**0.685**	**0.704**	**0.612**	**0.627**

and thus can influence algorithm performance. Since a graph structure is mainly determined by its adjacency matrix (Eq. 3), which is computed with Gaussian function and an alternative similarity function. We investigate different metric functions for computing the adjacency matrix. Here, we choose the Cosine similarity and the Euclidean distance for comparison. Note that we negate the Euclidean distance to make it consistent with a similarity. Specifically, we utilize the normalized inner product as Cosine similarity, which is defined in Eq. 3. Then, to construct the Euclidean distance-based graph, we remove the l_2-normalization process and replace the inner product by a l_2 norm:

$$W_{i,j} = \begin{cases} exp(-\frac{||\hat{x}_i, \hat{x}_j||_2}{2\sigma^2}), & if\, i \neq j \wedge \hat{x}_i \in \mathcal{N}_k(\hat{x}_j) \\ 0, & otherwise \end{cases}, \quad (11)$$

$i, j \in [1, N+M]$, where $N+M$ is the number of image features in $\hat{X}$. Note that $\hat{x}_i$ is the i-th element of feature set $\hat{X}$ and $\mathcal{N}_k(\hat{x}_j)$ is the set of the top k nearest neighbors of $\hat{x}_j$ in $\hat{X}$. Besides, σ is a predefined scale parameter for the Gaussian function and we follow [29] to set $\sigma^2 = 0.05$. As is shown in Table 5, the performances on two databases vary with different metric functions and Cosine similarity outperforms the Euclidean distance. Therefore, we report the other experimental results based on the Cosine similarity in this paper.

5.3 Individual Visualization

Previous analyses mainly focus on the average performance on a batch of test users, in this subsection, we delve into the performance of individual user. As is shown in Fig. 3, we present the performance of three methods (i.e., Base-PIAA, TAPP-PIAA w/o meta and TAPP-PIAA) on 37 test users from FLICKR-AES under the setting of 100-shot PIAA task. We observe that TAPP-PIAA (red) achieves the best performance on all test users. Besides, we can see that TAPP-PIAA w/o meta (yellow) outperforms our baseline model Base-PIAA (blue) on the vast majority of test users. This indicates that with the same pre-trained GIAA prior representation, our proposed TAPP-PIAA algorithm can learn individual aesthetic taste better than the baseline method based on fine-tuning. Moreover, we can also observe that TAPP-PIAA (red) outperforms TAPP-PIAA w/o meta (yellow) on all test users. This observation further proves the effectiveness of the meta-learning framework in PIAA task, which can enhance the learning algorithm's generalization ability when facing unseen users.

5.4 Value of k

In this subsection, we investigate the impact of k (in Eq. 3) on the performance of TAPP-PIAA. Specifically, the value of k nearest neighbors could influence the graph structure [29] and further affect the aesthetic knowledge propagation. In order to dig out the relationship between k and PIAA performance, we present the experimental results of TAPP-PIAA based on different k values on two databases. As is shown in Fig. 4, the value of k has a mild influence on PIAA results, which indicates that we should choose a proper k value for specific PIAA task. From the line chart we can observe that the results of k=30 and k=50 outperform the others on 10-shot and 100-shot PIAA tasks, respectively. Therefore, we set k=30 for 10-shot tasks and k=50 for 100-shot tasks when reporting the other results in this paper.

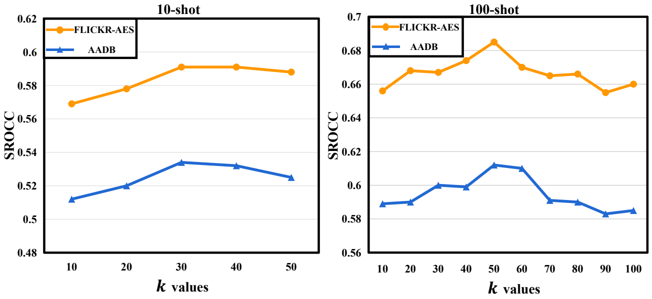

Figure 4: Experimental results with different k values.

6 DISCUSSION

6.1 Regression or Classification

In this paper, we introduce transductive label propagation mechanism into PIAA task for the first time, while it is often used for classification tasks [14, 18, 37]. In fact, PIAA task is different from classification tasks from the following aspects. First, in classification tasks, the label of each image is inherently fixed based on its objective properties, such as semantic. However, in PIAA task, the label of each image is assigned according to users' subjective preferences, which means an image could be annotated with different labels by different users. Second, in classification tasks, there are clear classification boundaries between different categories. However, in PIAA task, there is no clear boundary among labels (such as score 4 and score 5) and aesthetic features with different labels are highly correlated. Based on the above analyses, directly applying classification-based label propagation strategy with a classification loss into PIAA task is inappropriate. Therefore, in this paper, we modify the traditional classification-based label propagation (discrete labels) to a regression-based one (continuous label) with a MSE loss. To prove the above points, we compare the results of our regression-based TAPP-PIAA with classification-based TAPP-PIAA* in Table 6. We can observe that TAPP-PIAA outperforms TAPP-PIAA* on both FLICKR-AES and AADB and this indicates that the regression-based propagation strategy is more suitable for PIAA task.

6.2 Transductive Propagation Strategy

Except the propagation formula introduced in Subsection 3.5, we also explore another widely-adopted transductive propagation strategy [11, 14, 18] into PIAA task. Similarly, we denote the label matrix at t-th iteration as $f^t \in \mathbb{R}^{(N+M)\times 1}$ and define the iterative formula of this propagation strategy as:

$$f^{t+1} = \alpha \bar{A} f^t + (1-\alpha)Y, \tag{12}$$

where $\bar{A}$ is the normalized *adjacency matrix* and Y is the initial label matrix defined by Eq. 4. Note that $\alpha \in (0,1)$ controls the amount of transferred information from support labels. Here, we follow [11, 14, 18] and set α to 0.99. Similarly, the above formula can also lead to a closed-form solution [35] for f^*:

$$f^* = (1-\alpha)(I-\alpha\bar{A})^{-1}Y, \tag{13}$$

where f^* is the predicted label matrix of all support and query samples. Similar to Eq. 8, we use Eq. 13 to directly calculate the propagation results in practice. As is shown in Table 7, we compare the experimental results of two propagation strategies. We can observe that our TAPP-PIAA outperforms TAPP-PIAA† by 0.014 to 0.036 on FLICKR-AES and AADB. This reveals the fact that the direct label recovery can better maintain the stability of the initial label information than using a control parameter α in PIAA task.

Table 6: Comparison between classification-based propagation and regression-based propagation. Note that TAPP-PIAA* is implemented under classification-based propagation with cross-entropy loss.

Method	FLICKR-AES		AADB	
	10-shot	*100-shot*	*10-shot*	*100-shot*
TAPP-PIAA*	0.559	0.655	0.515	0.583
TAPP-PIAA	**0.591**	**0.685**	**0.534**	**0.612**

Table 7: Comparison between different transductive propagation strategies. Note that TAPP-PIAA† is implemented according to Eq. 13.

Method	FLICKR-AES		AADB	
	10-shot	*100-shot*	*10-shot*	*100-shot*
TAPP-PIAA†	0.565	0.662	0.520	0.576
TAPP-PIAA	**0.591**	**0.685**	**0.534**	**0.612**

7 CONCLUSION

In this work, we first-time propose using meta learning-based transductive label propagation algorithm into PIAA task and have proposed an algorithm named TAPP-PIAA for utilizing the potential value of unlabeled data when inferring personalized aesthetic preference. Further, we can also substitute the fine-tuning operation when deploying on personal data via personalized aesthetic preference embedding and graph inference. Here, to better fit the characteristics of PIAA task, we extend the existing classification-based propagation strategy into a regression-based one to learn to propagate aesthetic preferences by optimizing a standard mean square error loss. We also report the proposed algorithm performance on different PIAA benchmark databases and conduct cross-database validation. We further give explanations and analyses on regression or classification setting, propagation strategy, influence of hyper-parameter selection, backbone and metric function. Experimental results indicate that the proposed method can outperform the state-of-the-art methods.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 62176116, Grant 62073160, and Grant 71732003, and in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 20KJA520006.

REFERENCES

[1] Sungyong Baik, Myungsub Choi, Janghoon Choi, Heewon Kim, and Kyoung Mu Lee. 2020. Meta-learning with adaptive hyperparameters. In *Advances in Neural Information Processing Systems*, Vol. 33. 20755–20765.

[2] Sungyong Baik, Seokil Hong, and Kyoung Mu Lee. 2020. Learning to forget for meta-learning. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2379–2387.

[3] Haoxing Chen, Huaxiong Li, Yaohui Li, and Chunlin Chen. 2022. Shaping visual representations with attributes for few-shot recognition. *IEEE Signal Processing Letters* 29 (2022), 1397–1401.

[4] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. 2019. A closer look at few-shot classification. In *International Conference on Learning Representations*.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.

[6] Yubin Deng, Chen Change Loy, and Xiaoou Tang. 2017. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine* 34, 4 (2017), 80–106.

[7] Li Fe-Fei et al. 2003. A Bayesian approach to unsupervised one-shot learning of object categories. In *IEEE International Conference on Computer Vision*. 1134–1141.

[8] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*. 1126–1135.

[9] Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimiliano Pontil. 2018. Bilevel programming for hyperparameter optimization and meta-learning. In *International Conference on Machine Learning*. 1568–1577.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

[11] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. 2019. Label propagation for deep semi-supervised learning. In *IEEE Conference on Computer Vision and Pattern Recognition*. 5070–5079.

[12] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[13] Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. 2016. Photo aesthetics ranking network with attributes and content adaptation. In *European Conference on Computer Vision*. 662–679.

[14] Michalis Lazarou, Tania Stathaki, and Yannis Avrithis. 2021. Iterative label cleaning for transductive and semi-supervised few-shot learning. In *IEEE International Conference on Computer Vision*. 8751–8760.

[15] Leida Li, Hancheng Zhu, Sicheng Zhao, Guiguang Ding, and Weisi Lin. 2020. Personality-assisted multi-task learning for generic and personalized image aesthetics assessment. *IEEE Transactions on Image Processing* 29 (2020), 3898–3910.

[16] Wenbin Li, Lei Wang, Jinglin Xu, Jing Huo, Yang Gao, and Jiebo Luo. 2019. Revisiting local descriptor based image-to-class measure for few-shot learning. In *IEEE Conference on Computer Vision and Pattern Recognition*. 7260–7268.

[17] Yaohui Li, Huaxiong Li, Haoxing Chen, and Chunlin Chen. 2021. Local mutual metric network for few-shot image classification. In *Chinese Conference on Pattern Recognition and Computer Vision*. 443–454.

[18] Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, Eunho Yang, Sung Ju Hwang, and Yi Yang. 2018. Learning to propagate labels: Transductive propagation network for few-shot learnig. In *International Conference on Learning Representations*.

[19] Pei Lv, Jianqi Fan, Xixi Nie, Weiming Dong, Xiaoheng Jiang, Bing Zhou, Mingliang Xu, and Changsheng Xu. 2021. User-guided personalized image aesthetic assessment based on deep reinforcement learning. *IEEE Transactions on Multimedia* (2021).

[20] Pei Lv, Meng Wang, Yongbo Xu, Ze Peng, Junyi Sun, Shimei Su, Bing Zhou, and Mingliang Xu. 2018. USAR: An interactive user-specific aesthetic ranking framework for images. In *ACM International Conference on Multimedia*. 1328–1336.

[21] Jerome L Myers, Arnold D Well, and Robert F Lorch Jr. 2013. *Research design and statistical analysis*. Routledge.

[22] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2014. Collaborative filtering of color aesthetics. In *Workshop on Computational Aesthetics*. 33–40.

[23] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. (2017).

[24] Rosalind W Picard. 2000. *Affective computing*. MIT press.

[25] Jian Ren, Xiaohui Shen, Zhe Lin, Radomir Mech, and David J Foran. 2017. Personalized image aesthetics. In *IEEE International Conference on Computer Vision*. 638–647.

[26] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, Vol. 30.

[27] Vladimir N Vapnik. 1999. An overview of statistical learning theory. *IEEE transactions on neural networks* 10, 5 (1999), 988–999.

[28] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. *Advances in Neural Information Processing Systems* 29 (2016), 3630–3638.

[29] Fei Wang and Changshui Zhang. 2007. Label propagation through linear neighborhoods. *IEEE Transactions on Knowledge and Data Engineering* 20, 1 (2007), 55–67.

[30] Guolong Wang, Junchi Yan, and Zheng Qin. 2018. Collaborative and Attentive Learning for Personalized Image Aesthetic Assessment.. In *International Joint Conference on Artificial Intelligence*. 957–963.

[31] Weining Wang, Junjie Su, Lemin Li, Xiangmin Xu, and Jiebo Luo. 2019. Meta-learning perspective for personalized image aesthetics assessment. In *IEEE International Conference on Image Processing*. 1875–1879.

[32] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. 2020. Generalizing from a few examples: A survey on few-shot learning. *Comput. Surveys* 53, 3 (2020), 1–34.

[33] Yuzhe Yang, Liwu Xu, Leida Li, Nan Qie, Yaqian Li, Peng Zhang, and Yandong Guo. 2022. Personalized image aesthetics assessment with rich attributes. In *IEEE Conference on Computer Vision and Pattern Recognition*. 19861–19869.

[34] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. 2020. Few-shot learning via embedding adaptation with set-to-set functions. In *IEEE Conference on Computer Vision and Pattern Recognition*. 8808–8817.

[35] Dengyong Zhou, Olivier Bousquet, Thomas Lal, Jason Weston, and Bernhard Schölkopf. 2003. Learning with local and global consistency. *Advances in Neural Information Processing Systems* 16 (2003).

[36] Hancheng Zhu, Leida Li, Jinjian Wu, Sicheng Zhao, Guiguang Ding, and Guangming Shi. 2020. Personalized image aesthetics assessment via meta-learning with bilevel gradient optimization. *IEEE Transactions on Cybernetics* (2020), 1798–1811.

[37] Xiaojin Zhu. 2005. *Semi-supervised learning with graphs*. Carnegie Mellon University.