西安電子科技大學
XIDIAN UNIVERSITY

Personalized Image Aesthetics Assessment

Leida Li

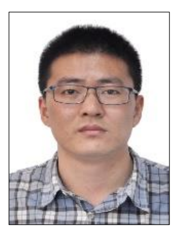
School of Artificial Intelligence
Xidian University

2020.05.27

Collaborators:


Hancheng Zhu
CUMT


Jinjian Wu
XDU


Sicheng Zhao
UC Berkeley


Guiguang Ding
Tsinghua


Guangming Shi
XDU


Weisi Lin
NTU

Contents

西安电子科技大学 XIDIAN UNIVERSITY

Contents

- **Introduction**
- Related Work
- Personality-assisted Multi-task Learning for Personalized Image Aesthetics Assessment
- Personalized Image Aesthetics Assessment via Meta-learning
- Conclusion

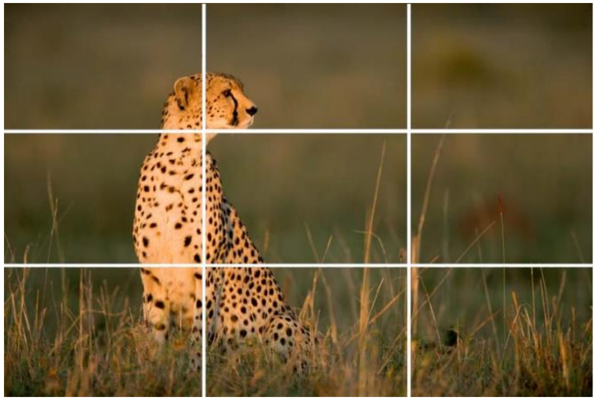
China has a proverb

爱美之心
人皆有之

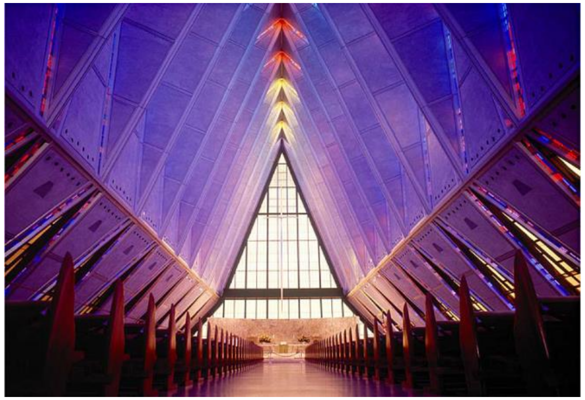
The love of beauty is common to all people

A natural question: ***how to judge aesthetics?***

- **Photography rules**



Rule of Thirds



Symmetry

Depth of Field

Color Harmony

K. Michal, et al., Leveraging expert feature knowledge for predicting image aesthetics. IEEE Trans. Image Process., 2018.

- **Photography rules**

https://baijiahao.baidu.com/s?id=1608952443864950129&wfr=spider&for=pc

- **Application scenarios**

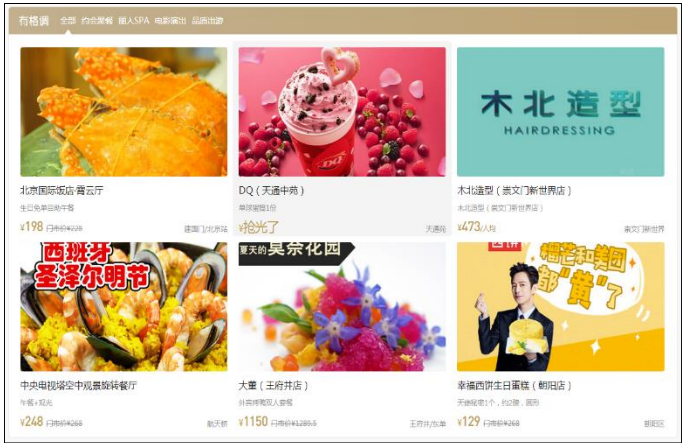
  - Advertisement

Alibaba's Luban system

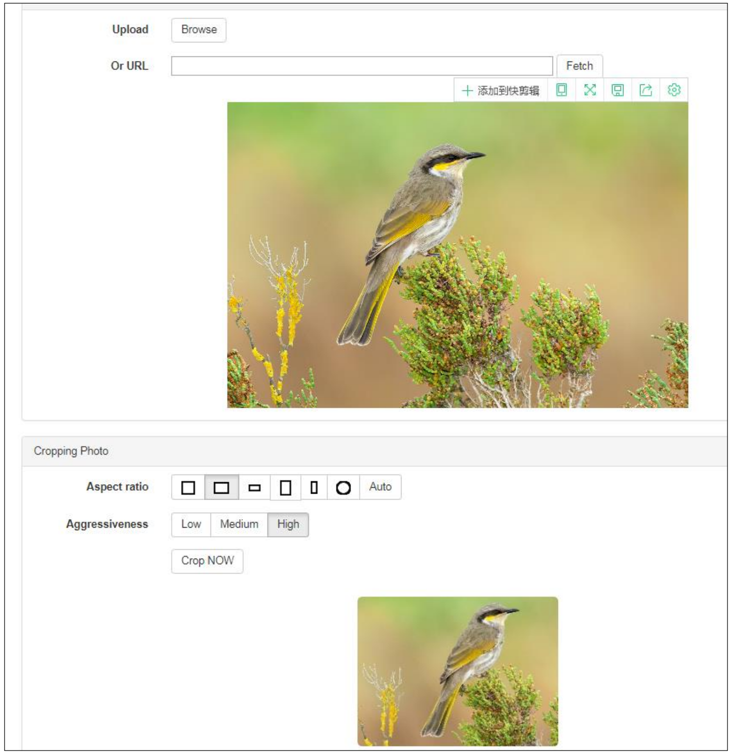
- Launched on 11/11/2016
- Designed 170 million posters
- Improved hit rate by 100%
- Equipped with **IAA engine**

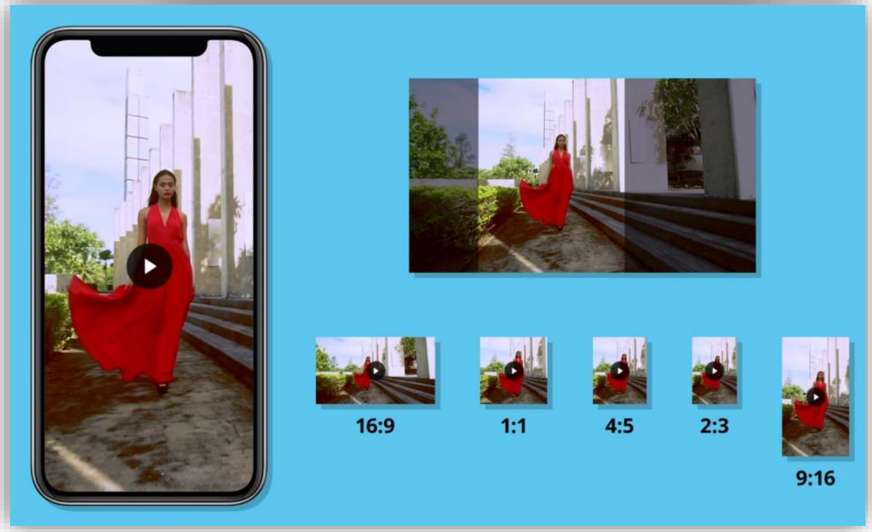
http://www.mgzxzs.com/tmtbzxjc/2288.html

- **Application scenarios**
  - **- Cover image selection**



- **Application scenarios**
  - **- Photo auto-cropping**



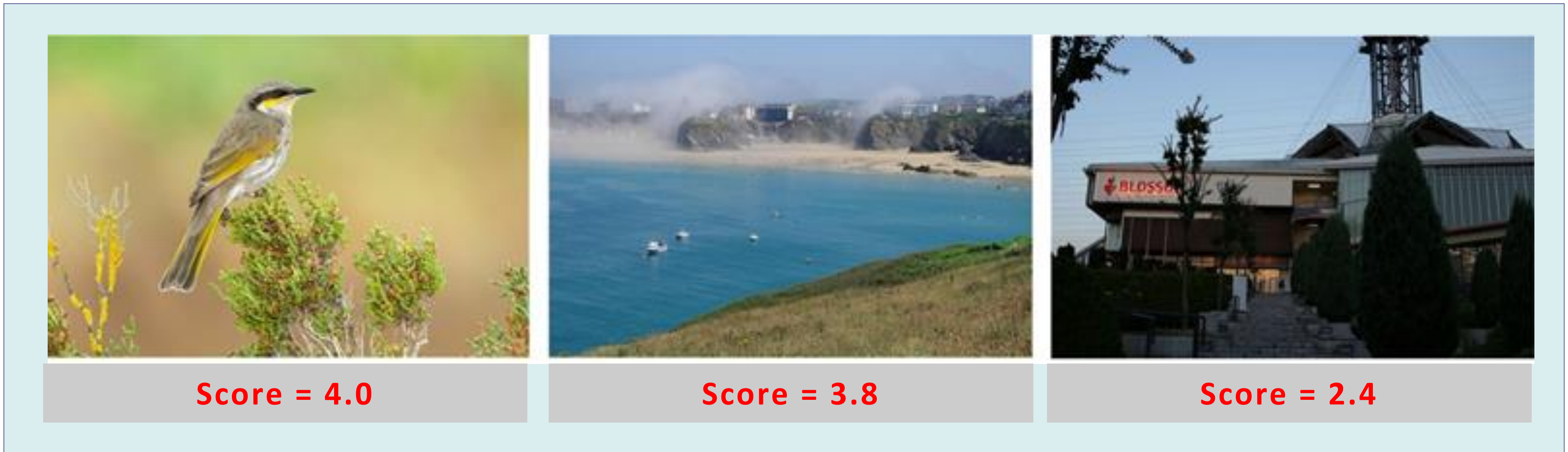


Contents

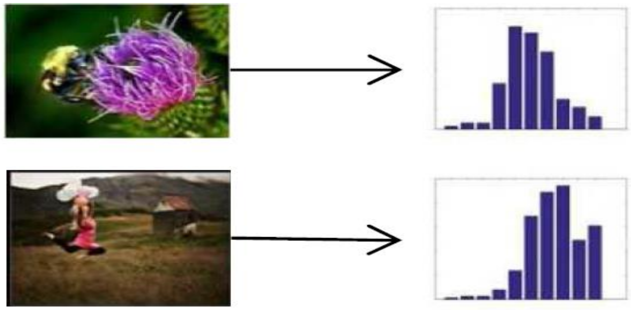
- Introduction
- **Related Work**
- Personality-assisted Multi-task Learning for Personalized Image Aesthetics Assessment
- Personalized Image Aesthetics Assessment via Meta-learning
- Conclusion

Generic Aesthetic Quality

- **Generic image aesthetics assessment (GIAA)**
  - Aesthetics classification
  - Aesthetics regression



- Aesthetics distribution



-H. Zeng, et al., A unified probabilistic formulation of image aesthetic assessment. IEEE Trans. Image Process., 2020
-Y. Deng, et al., Image aesthetic assessment an experimental survey. IEEE Signal Process. Mag., 2017.

Related Work

- **Conventional approaches with handcrafted features**
  - **Simple image features**
    - Colorfulness
    - Contrast
    - Brightness
  - **Image composition features**
    - Low depth of field
    - Salient object
    - Rule of thirds
  - **General-purpose features**
    - SIFT descriptors
    - Bag of visual words (BOV)
    - Fisher vector (FV)

High colorfulness

Low colorfulness

Good composition

Bad composition

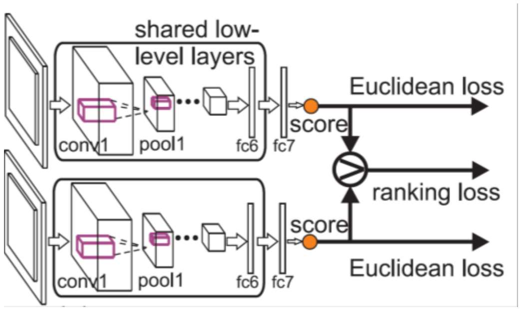
Extracted handcrafted features for image aesthetics assessment

-R. Datta, et al., Studying aesthetics in photographic images using a computational approach. ECCV 2006.
-X. Tang, et al., Content-based photo quality assessment. IEEE Trans. Multimedia, 2013.
-N. Murray, et al., AVA: A large-scale database for aesthetic visual analysis. CVPR 2012.

Related Work

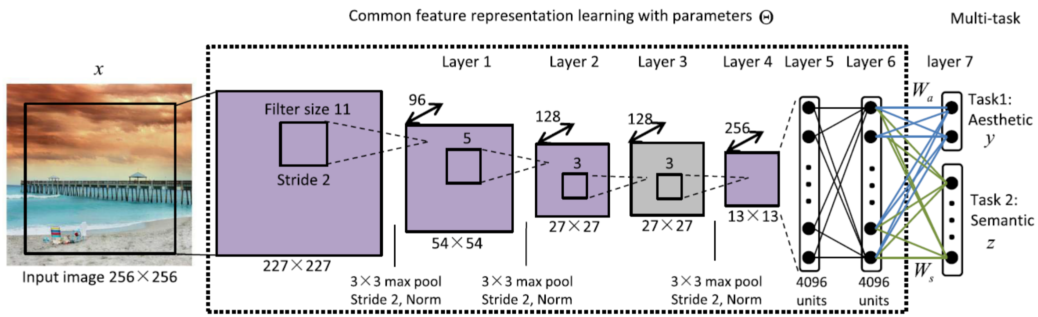
- **Deep-learning approaches**
  - **Ranking deep network**

Aesthetics attribute



Network architecture

  - **Multi-task deep network**



Significant progress has been achieved in GIAA.

China has another proverb

One man's meat is another man's poison

Personalized Aesthetic Quality

- **Personalized image aesthetics assessment (PIAA)**
  - People have different tastes on image aesthetics, depending on their **subjective preferences**.

Score = 4.0

Ratings = (3,3,4,5,5)

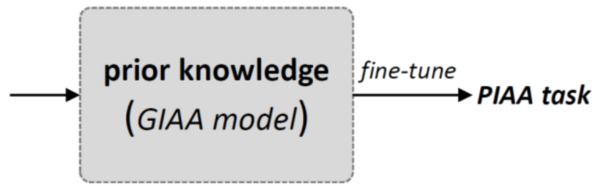
Score = 3.8

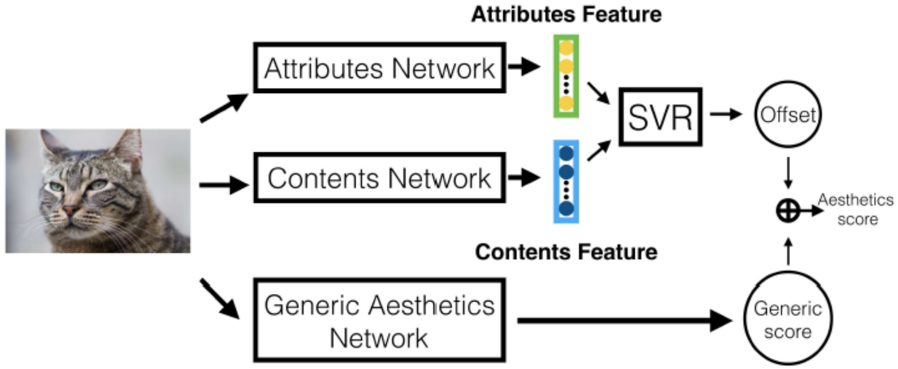
Ratings = (2,3,4,5,5)

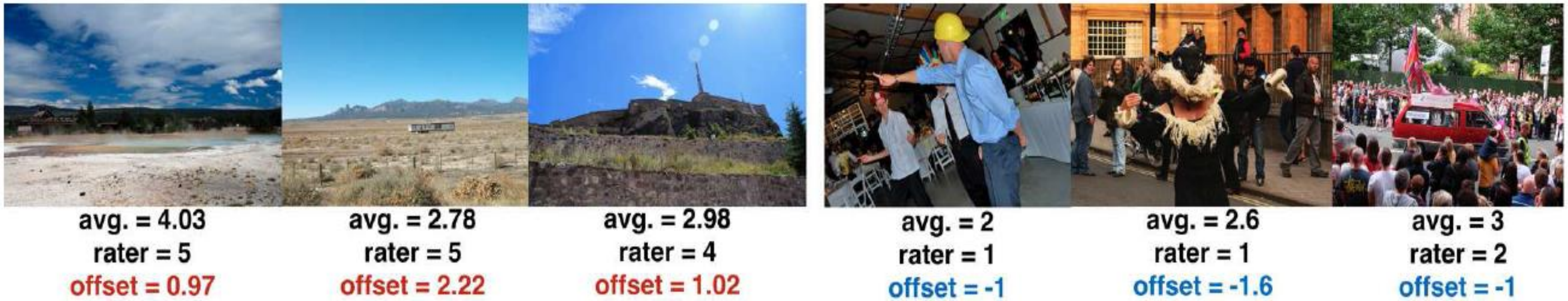
Score = 2.4

Ratings = (1,2,2,3,4)

- **Adapting from generic aesthetics**





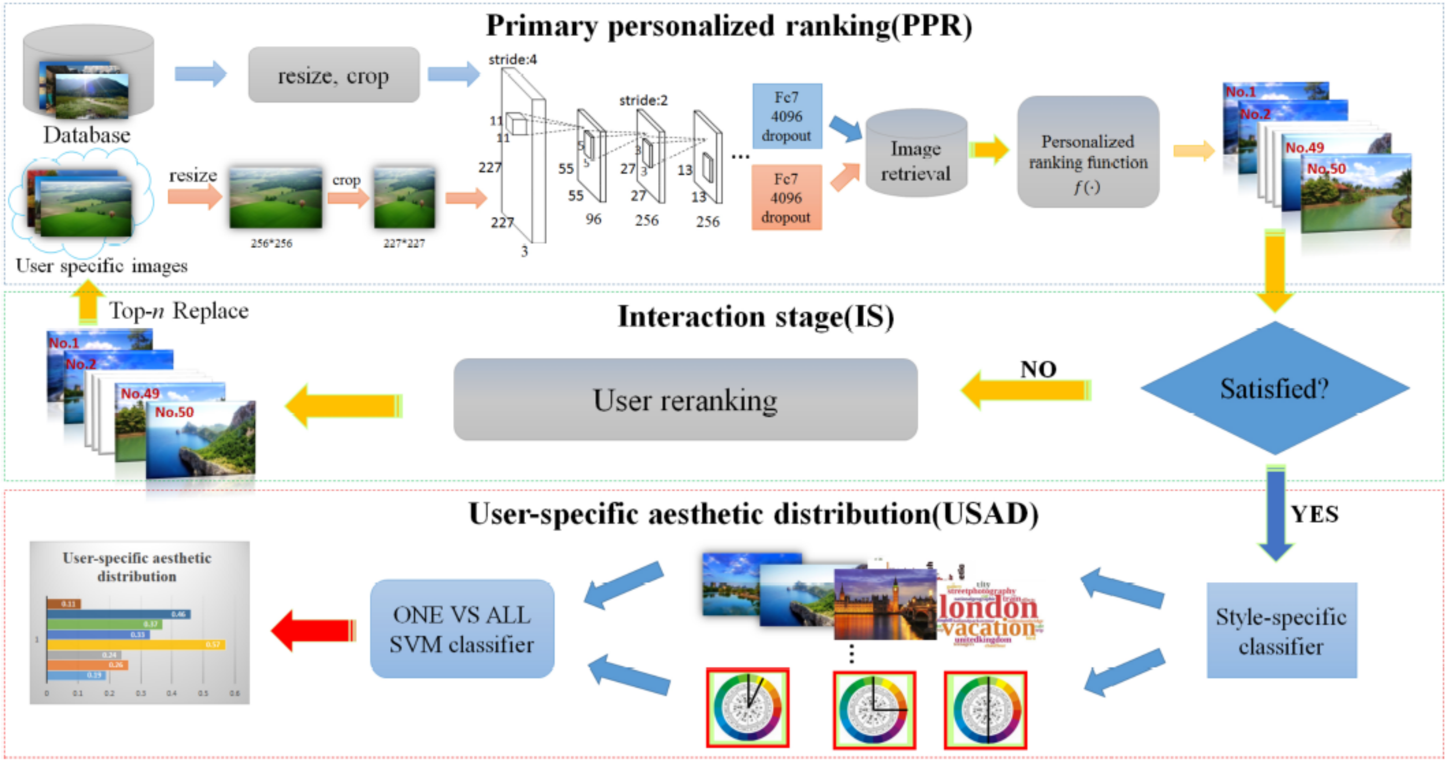


Personalized image aesthetics model with a residual-based model adaptation scheme.

J. Ren, et al., Personalized image aesthetics. ICCV 2017.

- **User interaction**

USAR : user-specific aesthetic ranking framework



User-friendly aesthetic ranking framework via deep neural network and a small amount of interaction.

P. Lv, et al., USAR: an interactive user-specific aesthetic ranking framework for images. ACM MM 2018.

Related Work

Challenges：

1. Existing works leverage **objective visual features** (e.g., contents and attributes) for modeling users' subjective aesthetic preferences. This may be insufficient, because the **subjective factors** (e.g., personality traits) in rating image aesthetics are not fully investigated.

2. The generic model learned from **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.

Contents

- Introduction
- Related Work
- **Personality-assisted Multi-task Learning for Personalized Image Aesthetics Assessment**
- Personalized Image Aesthetics Assessment via Meta-learning
- Conclusion

Personality-assisted Multi-task Learning for Generic and Personalized Image Aesthetics Assessment (L. Li, H. Zhu, et al., IEEE TIP, 2020)

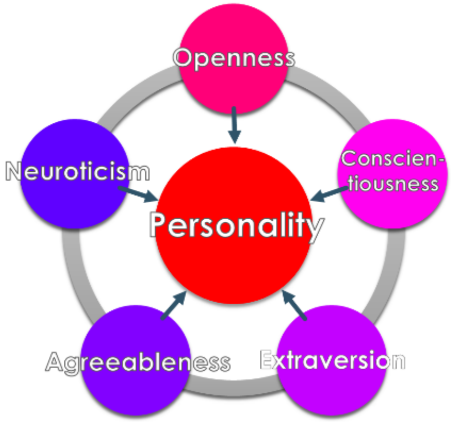
- As an important **subjective trait**, personality trait is believed as a key factor in modeling humans' subjective preferences.
- What is the relationship between **aesthetics assessment** and **personality prediction** from images?

Images liked by users with high extraversion — Images liked by users with low extraversion

-H. Zhu, L. Li, et al., Evaluating attributed personality traits from scene perception probability, Pattern Recogn Lett, 2018.
-S. C. Guntuku, et al., Who likes what, and why? insights into personality modeling based on image "likes", IEEE Trans Affect Comput, 2018.

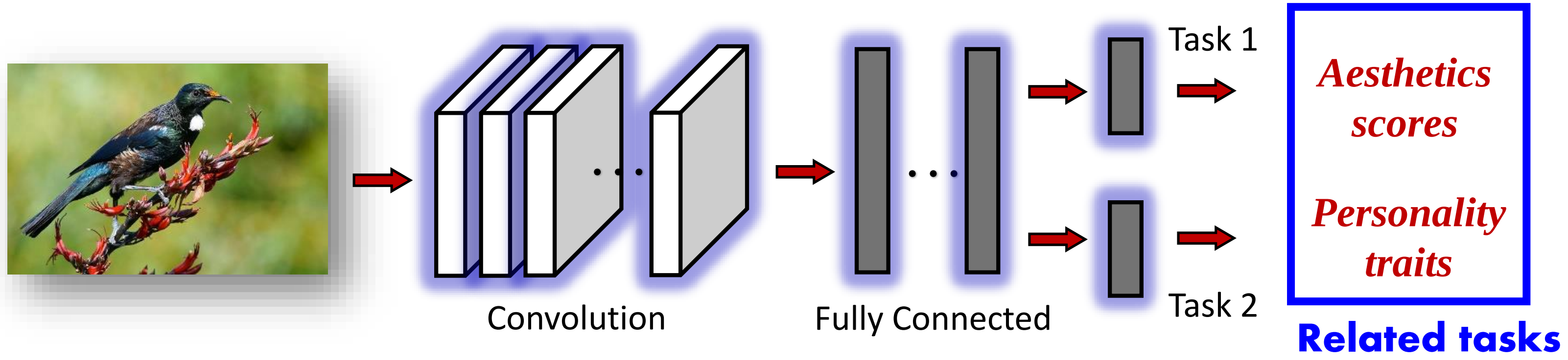
PA_IAA

- **Big-Five personality traits**
  - **Openness**: tendency to be open, curious, etc.
  - **Conscientiousness**: tendency to be responsible and reliable.
  - **Extraversion**: tendency to interact and spend time with others.
  - **Agreeableness**: tendency to be kind, generous, etc.
  - **Neuroticism**: tendency to be anxious, sensitive, etc.



- **Multi-task learning**

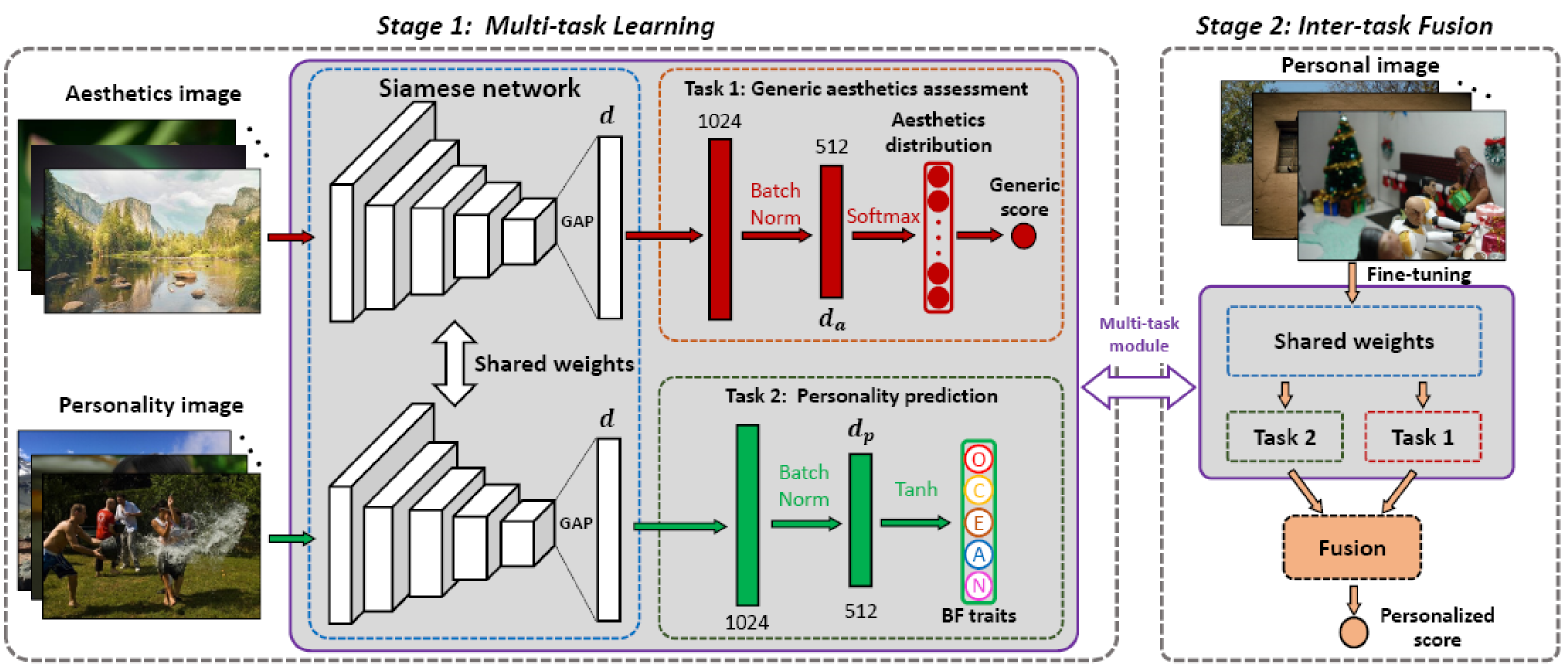
An effective way in capturing useful information contained in multiple **related tasks**, which can be used to improve the **generalization performance** of all tasks.



-B. Rammstedt, et al., Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German, J. Res. Pers. 2007.
-S. Ruder, CoRR, 2017. Online: http://arxiv.org/abs/1706.05098

PA_IAA

- A multi-task learning network with shared weights is proposed to predict the **aesthetics distribution** of an image and **Big-Five (BF) personality traits** of people who like the image
- To capture the common representation of image aesthetics and people's personality traits, a Siamese network is trained **using aesthetics data and personality data jointly**.
- **Inter-task fusion** is introduced to generate individual's personalized aesthetic scores.



PA_IAA

- Generic aesthetics training samples: $\{I_a^i; \boldsymbol{s}_a^i\}_{i=1}^{N_a}$, where $\boldsymbol{s}_a^i = \{s_{a_n}^i\}_{n=1}^{N}$; Aesthetic deep features: $\boldsymbol{d}_a$
  - Estimated aesthetic distribution:

$$\hat{s}_{a_n}^i = \frac{e^{\boldsymbol{W}_{a_n}^T \boldsymbol{d}_a}}{\sum_{j=1}^{N} e^{\boldsymbol{W}_{a_j}^T \boldsymbol{d}_a}}$$

  - Generic aesthetics loss function:

$$L_a = \frac{1}{N_a}\frac{1}{N}\sum_{i=1}^{N_a}\sum_{n=1}^{N}\left\|\hat{s}_{a_n}^i - s_{a_n}^i\right\|_2^2$$

- Personality training samples: $\{\{I_p^{um}\}_{m=1}^{M}, \{S_p^{ui}\}_{i=1}^{5}\}_{u=1}^{U}$, U is the number of users, M is the number of images liked by a user. Personality deep features: $\boldsymbol{d}_p$
  - Predicted personality distribution:

$$\widehat{\boldsymbol{S}}_p^{um} = \frac{e^{\boldsymbol{W}_p^T \boldsymbol{d}_p} - e^{-\boldsymbol{W}_p^T \boldsymbol{d}_p}}{e^{\boldsymbol{W}_p^T \boldsymbol{d}_p} + e^{-\boldsymbol{W}_p^T \boldsymbol{d}_p}}$$

  - Personality loss function:

$$L_p = \frac{1}{5}\frac{1}{U}\frac{1}{M}\sum_{u=1}^{U}\sum_{m=1}^{M}\sum_{i=1}^{5}\left\|\hat{S}_p^{umi} - S_p^{ui}\right\|_2^2$$

- Personalized aesthetics training samples: $\{I_b^i; s_b^i\}_{i=1}^{N_b}$;
  - Estimated personalized aesthetic score:

$$\hat{s}_b^i = \hat{s}_a^i + \boldsymbol{W}_b \hat{\boldsymbol{s}}_p^i$$

  - Personalized aesthetic loss function:

$$L_b = \frac{1}{N_b}\sum_{i=1}^{N_b}\left\|\hat{s}_b^i - s_b^i\right\|_2^2$$

Experiments

Databases:

- Aesthetics databases:
  - **GIAA: AVA** (250,000 images, 230,000 for training, 20,000 for testing)
  - **PIAA: FLICKR-AES** (40,000 images; 210 users, 173 for training, 37 for testing)
- Personality database:
  - **PsychoFlickr** (60,000 liked images of 300 users (200 images per user))

Hyper-parameters: weight decay of 1e-5, momentum of 0.9, batch size of 50, initial learning rate of 1e-4, drops to a factor of 0.9 every epoch, and total epoch of 50.

Criteria:

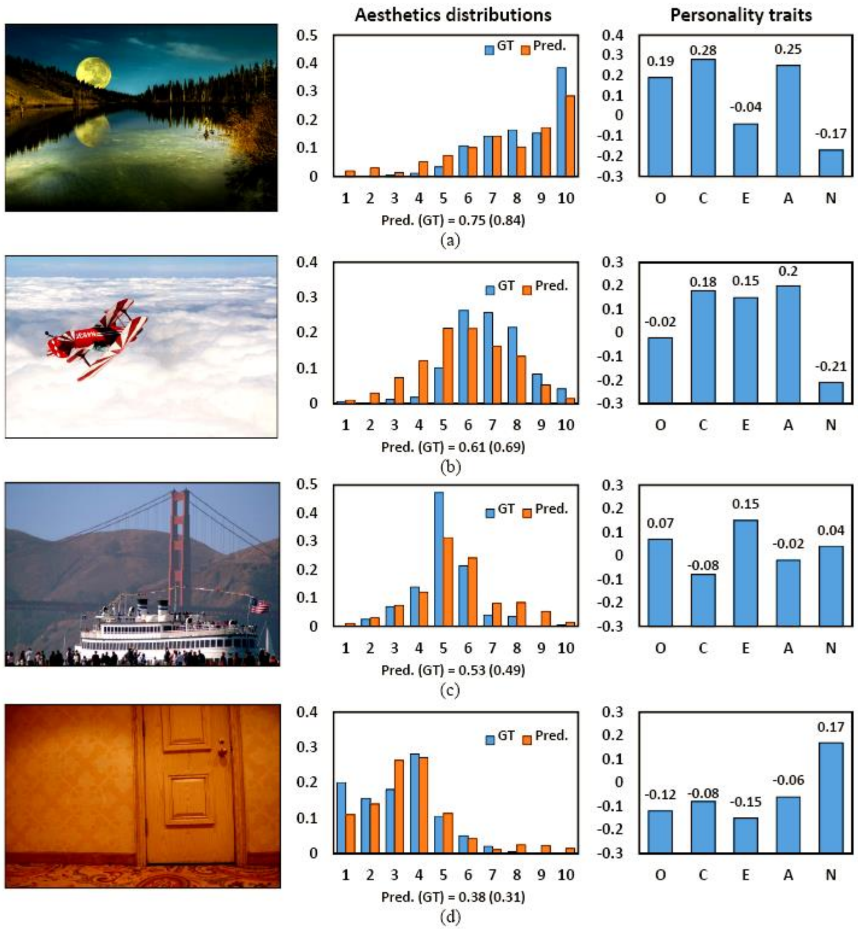
Classification: Overall Accuracy (ACC) ↑
Regression: Spearman Rank Order Correlation Coefficients (SROCC) ↑
Distribution: Earth Mover's Distance (EMD) ↓

-N. Murray, L. Marchesotti, and P. F., "AVA: a large-scale database for aesthetic visual analysis," ICCV 2012.
-S. Kong, X. Shen, Z. Lin, R. Mech, "Photo aesthetics ranking network with attributes and content adaptation," ECCV 2016.
-M. Cristani et al., "Unveiling the multimedia unconscious: implicit cognitive processes and multimedia content analysis," ACM MM 2013.

Experiments

Performance comparison on AVA database (Aesthetics classification)

Method	ACC(%)↑	SROCC↑	EMD↓
AVA handcrafted features [14]	68.0	-	-
RAPID [18]	74.5	-	
RAPID (improved version) [6]	75.4	-	-
DMA [19]	75.4	-	-
Wang *et al.* [20]	76.9	-	-
Kao *et al.* [27]	76.2	-	-
BDN [28]	78.1	-	-
Kao *et al.* [29]	79.1	-	-
Zhang *et al.* [30]	78.8	-	-
Schwarz *et al.* [31]	75.8	-	-
Kucer *et al.* [32]	81.9	-	-
ILGNet [33]	82.7	-	-
AlexNet_FT_Conf [2]	71.5	0.481	-
Reg+Rank+Att [2]	75.5	0.545	-
Reg+Rank+Cont [2]	73.4	0.541	-
Reg+Rank+Att+Cont [2]	77.3	0.558	-
USAR_PPR [22]	72.4	0.600	-
USAR_PAD [22]	77.7	0.545	-
USAR_PPR&PAD [22]	78.1	0.578	-
NIMA(VGG16) [36]	80.6	0.592	0.054
NIMA(Inception-v2) [36]	81.5	0.612	0.050
DenseNet121(aesthetics)	80.5	0.630	0.051
Inception-v3(aesthetics)	80.9	0.638	0.050
PA_IAA(DenseNet121)	82.9	0.666	0.049
PA_IAA(Inception-v3)	**83.7**	**0.677**	**0.047**



Aesthetics distribution prediction

Experiments

PIAA Database: FLICKR-AES

- 40,000 images, each image is labeled with aesthetic score by five different users.
  - Training: 35,263 images rated by 173 users
  - Testing: 4737 images rated by the other 37 users

Performance comparison on FLICKR-AES database

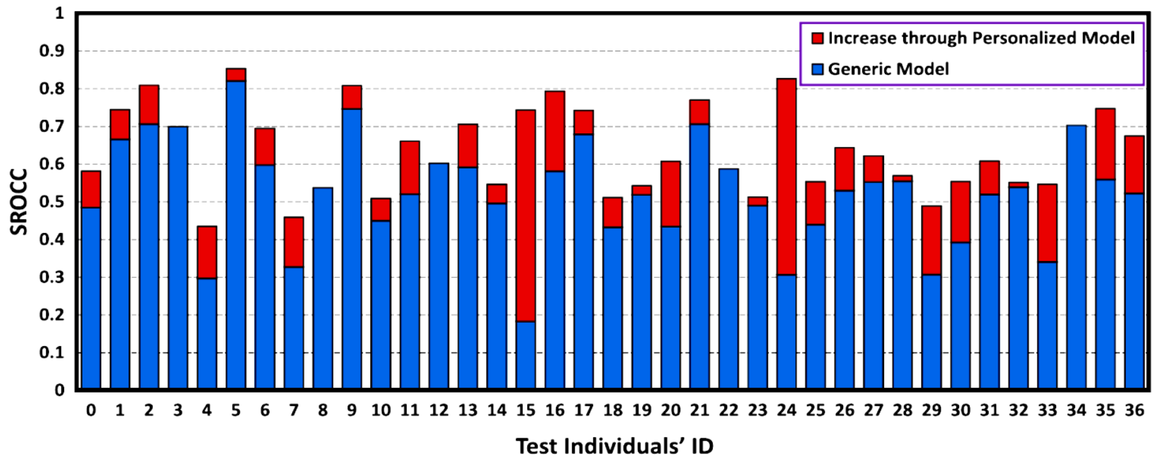
Method	10 images	100 images
FPMF (only attribute) [63]	0.511±0.004	0.516±0.003
FPMF (only content) [63]	0.512±0.002	0.516±0.010
FPMF (content and attribute) [63]	0.513±0.003	0.524±0.007
PAM (only attribute) [8]	0.518±0.003	0.539±0.013
PAM (only content) [8]	0.515±0.004	0.535±0.017
PAM (content and attribute) [8]	0.520±0.003	0.553±0.012
USAR_PPR [22]	0.521±0.002	0.544±0.007
USAR_PAD [22]	0.520±0.003	0.537±0.003
USAR_PPR&PAD [22]	0.525±0.004	0.552±0.015
MT_IAA	0.523±0.004	0.582±0.014
PA_IAA	**0.543±0.003**	**0.639±0.011**

PA_IAA outperforms MT_IAA by a large margin, which indicates that the personality prediction task of the proposed model has made a significant contribution to PIAA.

FLICKR-AES : J. Ren, X. Shen, Z. Lin, R. Mech, and D. J. Foran, Personalized image aesthetics, ICCV 2017.

Experiments

- **Performance comparison of test individuals**



- **Personality prediction performance on FLICKR-AES database**

Method	O	C	E	A	N
Segalin *et al.* [42]	0.354	0.535	0.625	0.476	0.613
Guntuku *et al.* [43]	0.398	0.552	0.679	0.525	0.636
DenseNet121(personality)	0.548	0.647	0.711	0.638	0.698
Inception-v3(personality)	0.536	0.654	0.703	0.651	0.709
PA_IAA(DenseNet121)	**0.567**	0.659	**0.722**	0.646	0.708
PA_IAA(Inception-v3)	0.555	**0.668**	0.715	**0.662**	**0.717**

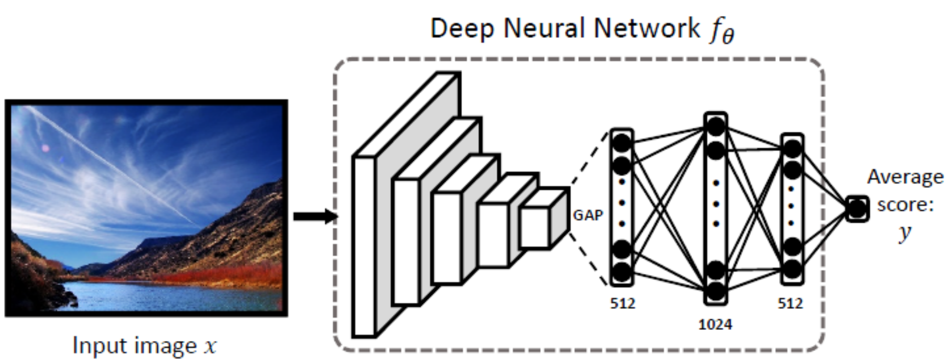
Multi-task learning module also contributes helpful information for personality prediction.

西安电子科技大学
XIDIAN UNIVERSITY

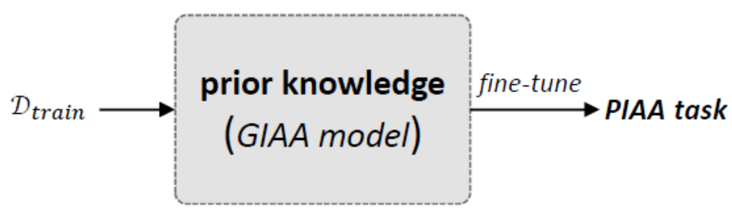
Contents

Personalized Image Aesthetics Assessment via Meta-learning with Bi-level Gradient Optimization (H. Zhu, L. Li, et al., IEEE TCYB, 2020)

- Existing PIAA models: fine-tuning **generic** image aesthetics assessment (GIAA) model as prior knowledge, which is not sufficient.
- Meta-learning is adopted to learn the **shared prior knowledge** among different people.
- **Fast adaptation** to unknown user for PIAA using very few sample images.



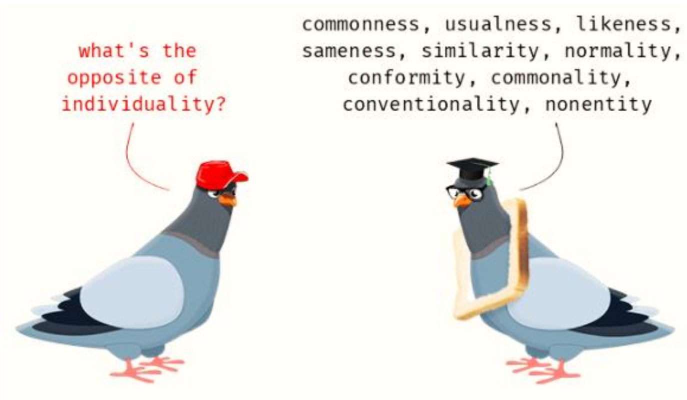
Network structure of GIAA model

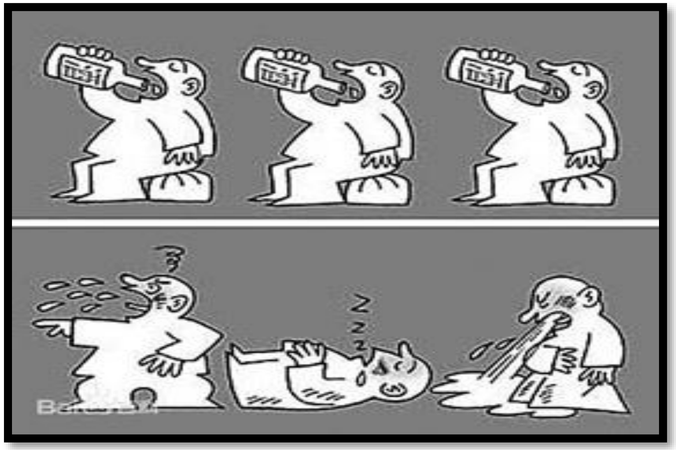


Fine-tune GIAA model to PIAA task

PIAA using "average aesthetics" based GIAA model as prior knowledge

- **Commonality and Individuality**





Commonality-based Individuality

High Aesthetics

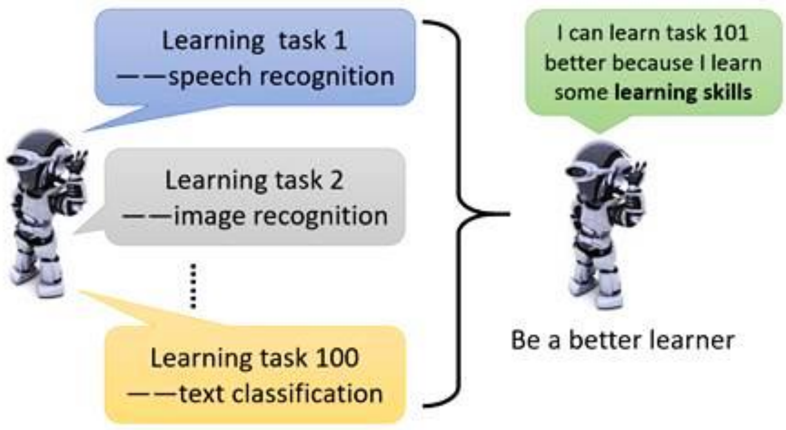
Low Aesthetics

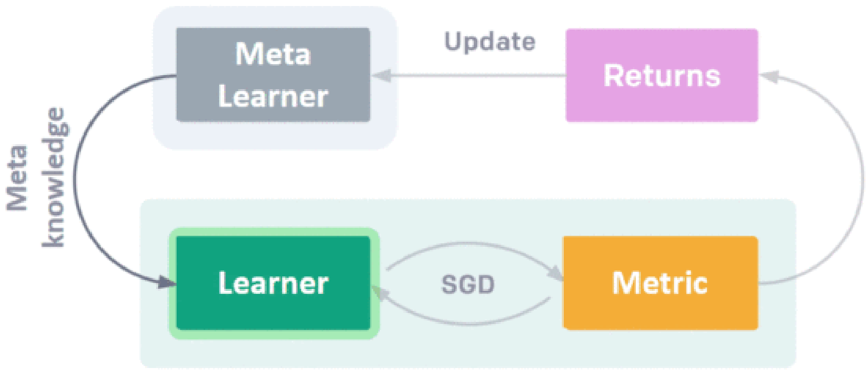
High agreement in binary classification

Share aesthetic knowledge

Meta-Learning (learn to learn)

- **Knowledge-driven** machine learning framework, which is able to extract meta knowledge from many related tasks.
- Deep meta learner can **fast adapt** to a new task with **limited training data**.



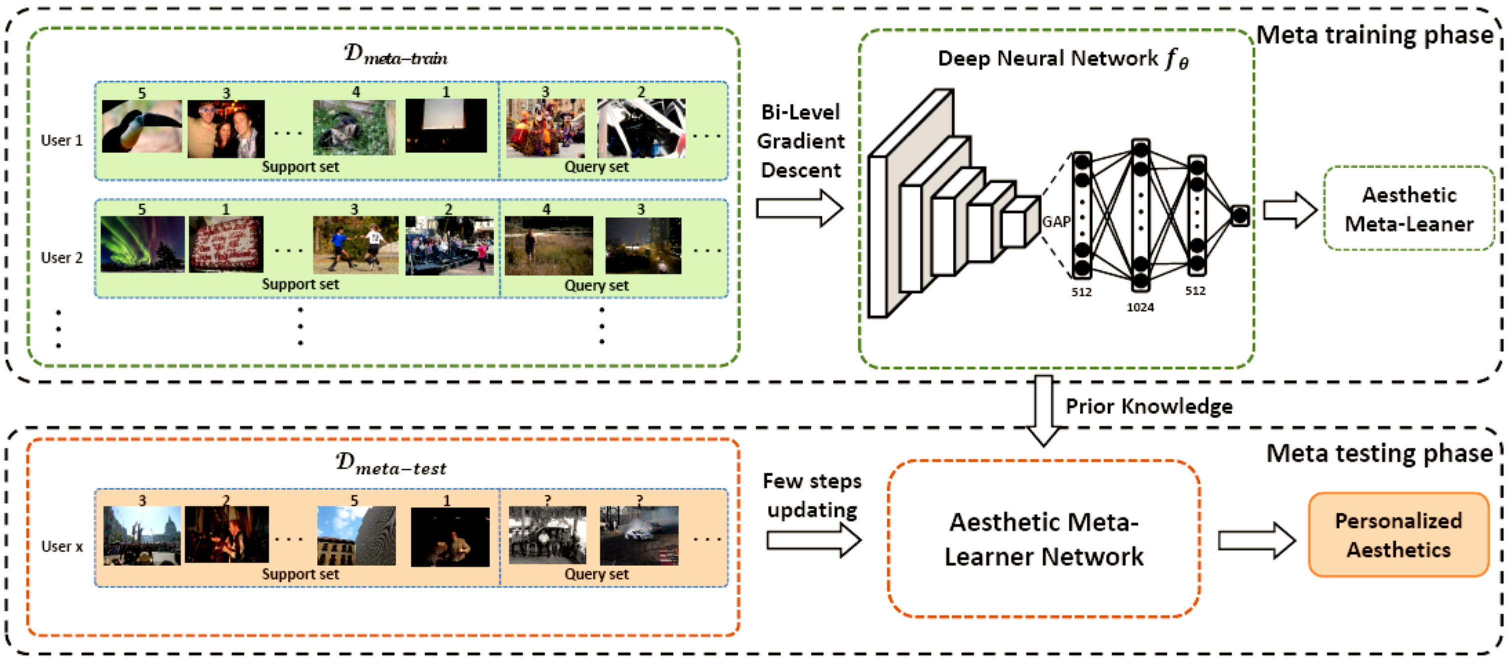


Vanschoren et al. Meta-Learning: A Survey. *arXiv, 2018.*

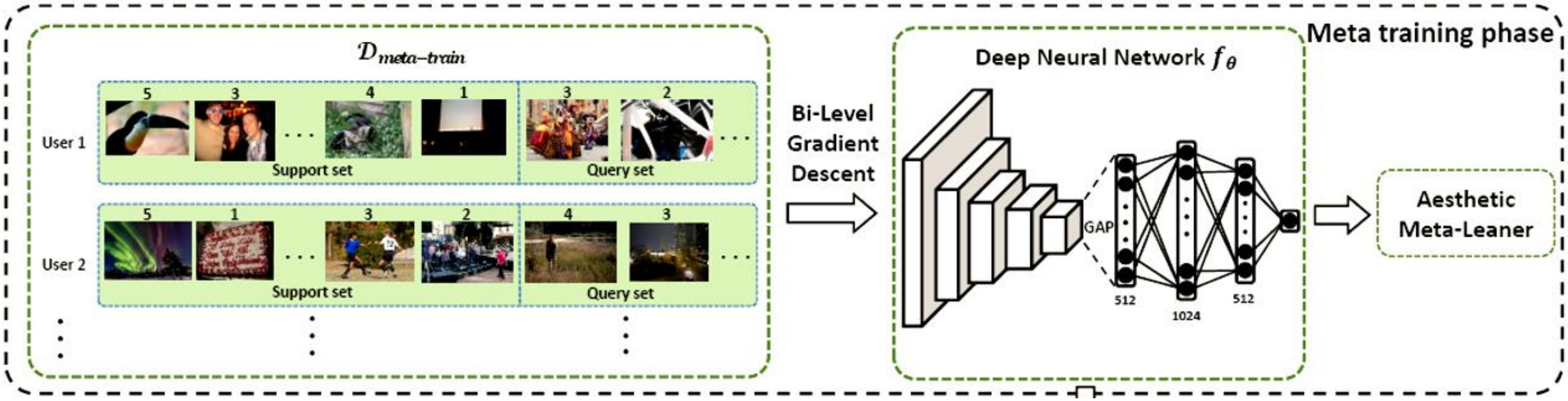
BLG-PIAA

Key idea:

- Meta-learning is adopted to learn the **shared prior knowledge** among different people in judge aesthetics.
- **Fast adaptation** to unknown user for PIAA using very few sample images.



BLG-PIAA



- Meta-training set: $\{\mathcal{D}_{tr_s}^{\tau_i}, \mathcal{D}_{tr_q}^{\tau_i}\}_{i=1}^{N}$;
- Deep neural network: f_θ;
- Loss function: $L_\tau = \frac{1}{M}\sum_{i=1}^{M}\|\hat{y}_i - y_i\|_2^2$;
- Gradient of network parameters: $g_\theta = \nabla_\theta L_\tau f_\theta$;
- First moment of $g_{\theta^{(s)}}$: $m_{\theta^{(s)}} = \mu_1 m_{\theta^{(s-1)}} + (1-\mu_1) g_{\theta^{(s)}}$;
- Second moment of $g_{\theta^{(s)}}$: $v_{\theta^{(s)}} = \mu_1 v_{\theta^{(s-1)}} + (1-\mu_1) {g^2}_{\theta^{(s)}}$;
- $Adam(L_\tau, \theta)$: $\theta \leftarrow \theta - \alpha \sum_{s=1}^{S} \frac{m_{\theta^{(s)}}}{\sqrt{v_{\theta^{(s)}}}+\epsilon}$;

```
Initialize model parameters θ: pre-trained on Imagenet;
/* meta-training phase */
for iteration = 1, 2, ... do
    Sample a batch of k tasks in D^{p(τ)}_{meta-train};
    for i = 1, 2, ..., k do
        /* first level computing */
        Compute θ'_i = Adam(L_{τ_i}, θ) for S steps on D^{τ_i}_{tr_s};
        /* second level computing */
        Compute θ_i = Adam(L_{τ_i}, θ'_i) for S steps on D^{τ_i}_{tr_q};
    end for
    update θ ← θ − β Σ_{i=1}^{k} (θ − θ_i);
end for
```

Performances on FLICKR-AES database.

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) [67]	0.511±0.004	0.516±0.003
FPMF (only content) [67]	0.512±0.002	0.516±0.010
FPMF (content and attribute) [67]	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 [42]	0.521±0.002	0.544±0.007
USAR_PAD [42]	0.520±0.003	0.537±0.003
USAR_PPR&PAD [42]	0.525±0.004	0.552±0.015
BA-PIAA	0.524±0.004	0.583±0.014
BLG-PIAA	**0.561±0.005**	**0.669±0.013**

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

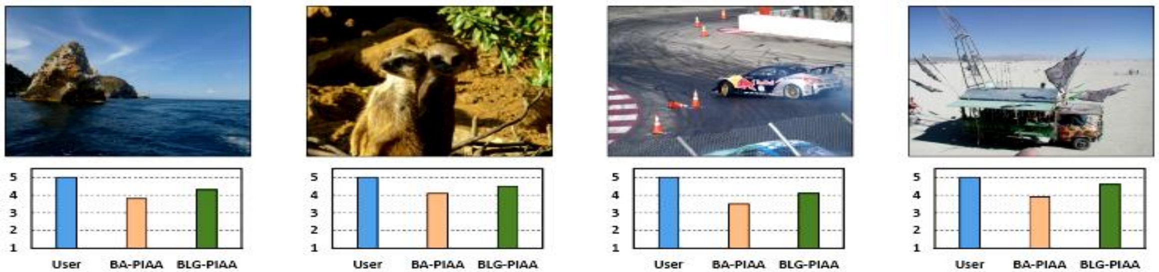
***BA-PIAA: baseline PIAA method based on "average aesthetics"**

BLG-PIAA further achieves **3.7%** and **8.6%** performance improvement when 10 and 100 images are used for training, respectively

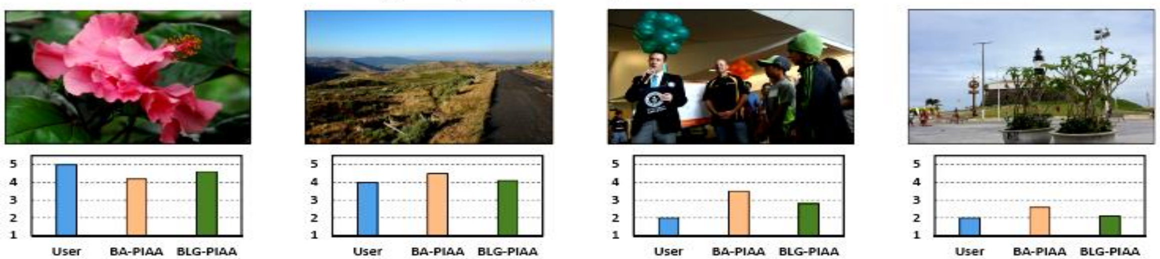
西安电子科技大学
XIDIAN UNIVERSITY

Visual results

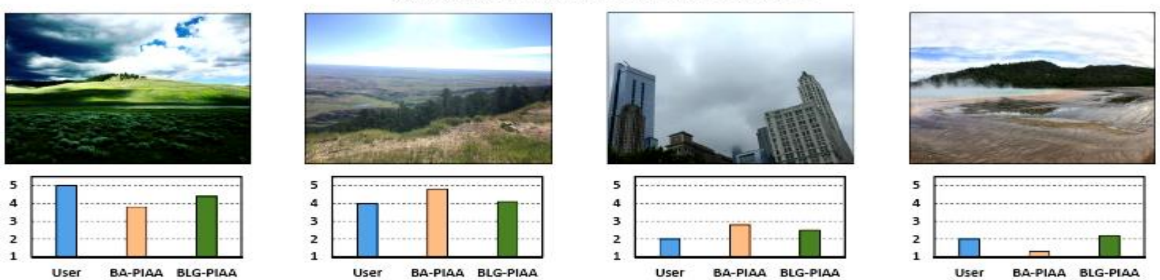
Visual Analysis



(a) Example images rated by a user from FLICKR-AES



(b) Example images rated by a user from AADB

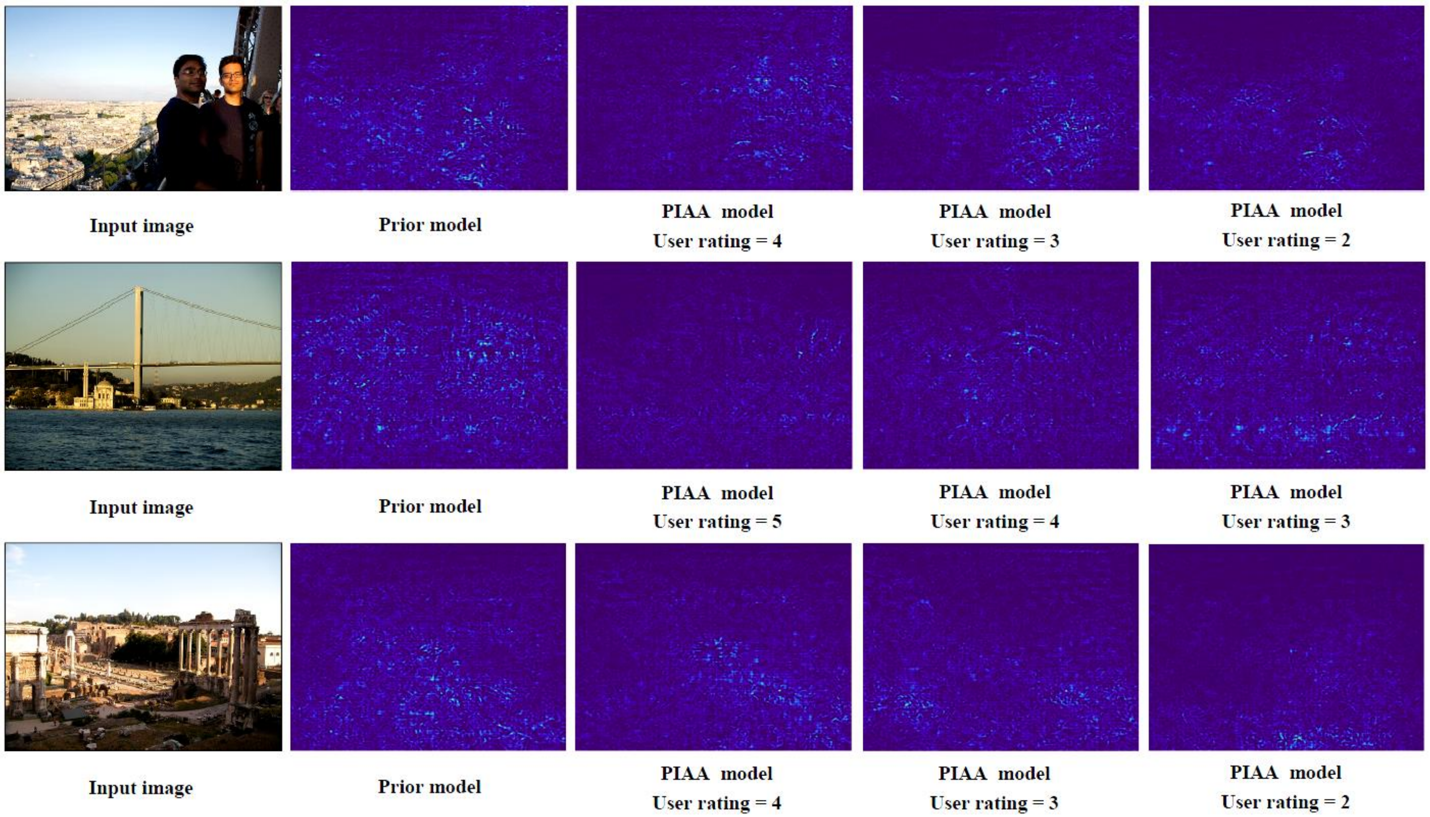


(c) Example images rated by a user from REAL-CUR

BLG-PIAA predicts user's aesthetic score more accurately than BA-PIAA.

Visual results

Visual Analysis



- The gradients map of user's PIAA model are more concentrated in **salient regions** than that of prior model.
- Users with different aesthetic ratings on an image have different **areas of interest**.

https://github.com/sar-gupta/convisualize_nb

Contents

Conclusion

- **Personality**, a key factor in subjective traits, has been utilized for personalized image aesthetics assessment under a multi-task learning framework. Personality data and aesthetics data were jointly used to learn the common features for predicting both aesthetics distribution and personality. Inter-task fusion was introduced to learn the influence of personality traits in individuals' aesthetic preferences on images.
- **Meta-learning** has been utilized to learn the shared prior knowledge in aesthetics assessment. By treating each individual's aesthetic assessment as a separate task, prior knowledge was learned, based on which PIAA was achieved by fast adaptation using only small samples.
- **User portrait** facilitates deeper understanding of user's aesthetic preference, personal interest, which is expected to benefits PIAA. This can be done using social data.

Thank you!