PERSONALITY DRIVEN MULTI-TASK LEARNING FOR IMAGE AESTHETIC ASSESSMENT

Leida Li[1,2], Hancheng Zhu[*2], Sicheng Zhao[3], Guiguang Ding[4], Hongyan Jiang[5], Allen Tan[6]*

[1]School of Artificial Intelligence, Xidian University, Xi'an 710071, China
[2]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
[3]Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley 94710, USA
[4]School of Software, Tsinghua University, Beijing 100084, China
[5]School of Management, China University of Mining and Technology, Xuzhou 221116, China
[6]Tencent Media Lab, Tencent, Shenzhen 518000, China

ABSTRACT

With the prevalence of convolutional neural networks (C-NNs), assessing the aesthetics of an image has gained great advances recently. Individual users often have different aesthetic preferences on images, which we believe are mainly affected by their personality traits. However, most of the current aesthetics models predict a generic aesthetic score based on handcrafted and/or learned feature representations, which are unified and thus cannot reflect the individual differences during image aesthetic rating. In this paper, we propose an end-to-end personality driven multi-task deep learning model to address this problem. Firstly, both image aesthetics and personality traits are learned from the proposed multi-task model. Then the personality features are employed to modulate the aesthetics features, producing the optimal generic image aesthetics scores. The experimental results on two public databases show that the proposed method is superior to the state-of-the-art approaches.

***Index Terms*—** Image aesthetic assessment, personality traits, multi-task deep learning, convolutional neural network

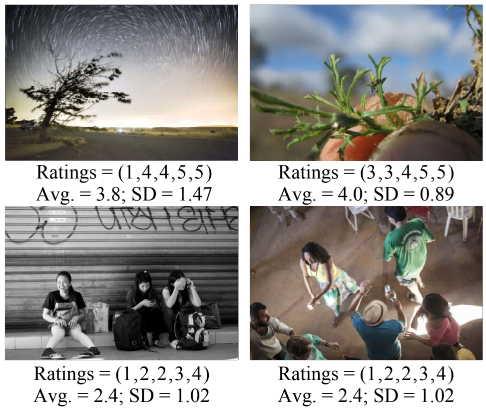

Fig. 1. Four example images associated with aesthetic scores rated by five different users from AADB database. The aesthetic scores are rated from 1 to 5.

1. INTRODUCTION

Image aesthetic assessment (IAA) aims at measuring people's aesthetic perception of images through photographic rules [1]. The automatic assessing of image aesthetics has many applications, including photo recommendation [2], photo cropping [3], image retrieval [4] and photo ranking [5]. Judging image aesthetics needs high-level understanding of the photographic attributes, which is extremely challenging. In recent years, many data-driven approaches have been proposed for IAA [5, 6, 7, 8], which attempt to learn the generic image aesthetic quality rated by an "average" user. However, users' aesthetic ratings on an image may vary significantly depending on their unique visual preferences [9]. For example, Figure 1 shows four images and the associated aesthetic scores rated by five different users from the Abstract Aesthetics and Attribute Database (AADB) [5]. The average (Avg.) and standard deviation (SD) of aesthetic scores are also shown. As illustrated in Figure 1, the aesthetic scores of an image differ significantly among different users whether the image has high or low average aesthetic score. This indicates that image aesthetics is affected by not only image content but also users' visual preferences. Therefore, it is necessary to learn the generic model for image aesthetics by taking users' preferences into account.

*Corresponding author (e-mail: zhuhancheng@cumt.edu.cn). This work was supported by the Natural Science Foundation of Jiangsu Province (BK20181354), National Natural Science Foundation of China (61771473 and 61379143), the Six Talent Peaks High-level Talents in Jiangsu Province (XYDXX-063) and the Qing Lan Project.

978-1-5386-9552-4/19/$31.00 ©2019 IEEE
DOI 10.1109/ICME.2019.00081

(a) (b)

Fig. 2. Two example images liked by users with different personality traits from the PsychoFlickr database: (a) an image liked by a user with high agreeableness; (b) an image liked by a user with high neuroticism.

Studies have shown that users' preferences on images are mainly influenced by their personality traits [10, 11]. For example, Figure 2 shows two images liked by users with different personality traits from the PsychoFlickr database [12]. As shown in Figure 2 (a), affable users tend to prefer image with natural scenes (mountains and lakes, etc.). Figure 2 (b) is an image liked by a user with high neuroticism, which manifests that neurotic users usually prefer image with dark and closed scenes. This indicates that the aesthetic preferences on images are intimately related to users' personality traits. Thus, it is reasonable to assess the generic image aesthetics by leveraging users' personality traits for capturing aesthetic differences on images.

In this paper, we address the problem of automatically assessing generic image aesthetics with the help of personality traits. An end-to-end personality driven multi-task learning model is designed to learn both the image aesthetic quality and the personality traits of users who like this image. In our multi-task model, a common representation for the two tasks can be learned in parallel with shared layers. Based on the relationship between personality traits and image aesthetics, we introduce an inter-task correlation learning for further improving the performance of image aesthetic assessment. The main contributions of our work are three-fold. (1) We propose a multi-task learning method for generic image aesthetic assessment, which outperforms previous works on two public aesthetic databases. (2) Our model can learn both generic image aesthetics and the personality traits of users who like this image simultaneously. (3) The proposed approach can automatically learn the inter-relationship between personality traits and image aesthetics. It can learn the discriminative representations of aesthetic discrepancy in modeling generic image aesthetics by an inter-task correlation learning.

2. RELATED WORKS

Image Aesthetic Assessment. Earlier works on image aesthetic assessment have focused on mapping handcrafted features to generic image aesthetic ratings [6, 7]. In [6], the low-level and high-level visual features were combined to train a SVM model for binary aesthetics classification. Tang *et al.* [7] proposed to extract multiple regional features in different ways according to image content for modeling the image aesthetics. In recent years, a large-scale Aesthetics Visual Analysis (AVA) database [8], which contains more than 250,000 labeled aesthetic images, was released. Besides, with the powerful feature representation of deep Convolutional Neural Networks (CNNs) [13], many works have been done to learn deep models for image aesthetic assessment [4, 5, 14, 15]. To the authors' best knowledge, the first attempt to train a CNN model for assessing image aesthetics was presented in [4], where an AlexNet-like framework was adopted. Lu *et al.* [14] proposed a Deep Multi-patch Aggregation (DMA) network for modeling the aesthetic quality by using multiple patches from an image. Wang *et al.* [15] developed a CNN-based method by incorporating seven scene categories for binary aesthetics classification. Instead of binary classification, Kong *et al.* [5] proposed to learn a Siamese network with adaptive image attribute and content information for image aesthetic regression and ranking. While these approaches have achieved notable success, the subjective factors of users in modeling image aesthetics are largely neglected.

Personality Computing. Personality computing aims at capturing users' stable preferences by exploring their observable behavioral cues [16]. With the prevalence of mobile internet, more and more images are uploaded in social networks, which greatly promotes social media-based personality computing [17]. In [12], 60,000 liked images of 300 Flickr individual users (200 images per user) were collected in the PsychoFlickr database. The Big-Five (BF) personality traits of each user are calculated by BIF-10 [18] questionnaire. In [10], Zhu *et al.* proposed a personality computing method based on image scene perception, which showed that individual user's preferences were affected by the statistics of image scenes. In [11], the local preferences of individuals on images were learned by a Weakly Supervised Dual Convolutional Network (WSDCN) for personality prediction. In the WSDCN model, user-level personality traits were used as the image-level personality labels for training. Therefore, the image-level personality traits can reflect user's aesthetic preferences on the image.

Multi-task Learning. Multi-task learning is an effective approach for improving the generalization performance by learning multiple related tasks with shared information simultaneously [19]. With the development of deep neural network, several deep network-based multi-task learning methods have been employed to address the problem of IAA. Kao *et al.* [20] divided the images into three categories (i.e., scene, object, texture), which were further used to train a classification model for image aesthetic assessment. In [21], Kao *et al.* found that the images' semantic content had an impact on people's aesthetic ratings, and proposed learning image aesthetics by incorporating the semantic information. The multi-task learning has shown to be an effective strategy to capture common

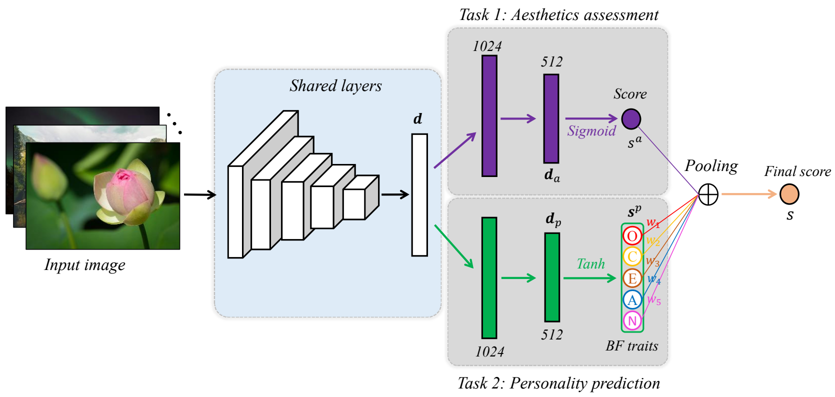

Fig. 3. The architecture of the proposed multi-task learning model for generic image aesthetic assessment.

features for related tasks simultaneously. Inspired by this, we develop a multi-task deep network to learn the relationship between image aesthetics and image personality attributes.

3. PROPOSED APPROACH

In this section, we propose to evaluate the generic image aesthetics with the help of personality traits. The architecture of the proposed approach is shown in Figure 3. Firstly, we design a multi-task learning network to predict both image aesthetic scores and BF personality traits simultaneously. In order to capture the common information for aesthetic task and personality task, the multi-task network with shared layers is then alternately trained based on two domain data (aesthetics and personality). Finally, the inter-task correlation learning is introduced for predicting the final generic aesthetic scores. The remainder of this section discusses the proposed multi-task learning network in detail.

Aesthetic Task Network. The DenseNet [22] is used as the basic model of our multi-task network, where the main task is generic aesthetic prediction and the auxiliary task is personality prediction. Since image aesthetic prediction is a regression problem, we replace the last output layer with two Fully Connected (FC) layers, which contain 1,024 nodes and 512 nodes, respectively. Following the last FC layer, a Sigmoid operator is used to generate the estimated aesthetic score $\hat{s}_i^a$, which is defined as

$$\hat{s}_i^a = \frac{1}{1 + e^{-\boldsymbol{W}_a^{\mathrm{T}} \boldsymbol{d}_a}}, \tag{1}$$

where $\boldsymbol{W}_a$ denotes the weight of estimated aesthetic score $\hat{s}_i^a$ from the last FC layer $\boldsymbol{d}_a$. We employ the Euclidean loss to learn the generic aesthetic task, and it is formulated as

$$L_a = \frac{1}{N_a} \sum_{i=1}^{N_a} \|\hat{s}_i^a - s_i^a\|_2^2, \tag{2}$$

where N_a is the number of images for training and s_i^a is the average aesthetic score rated by multiple users for the i-th image.

Personality Task Network. As mentioned above, the auxiliary task is to predict the personality attributes associated with an image, indicating the five personality traits of users who like this image. In order to learn the common features for both image aesthetics and personality traits, we introduce a personality task in parallel with the aesthetic task by using a shared representation. The personality network utilizes another two FC layers for predicting the BF personality scores. Following the last FC layer, a Tanh operator is used to produce five personality scores $\hat{\boldsymbol{s}}_i^p = \{\hat{s}_{i,j}^p\}_{j=1}^5$, which is defined as

$$\hat{\boldsymbol{s}}_i^p = \frac{e^{\boldsymbol{W}_p^{\mathrm{T}} \boldsymbol{d}_p} - e^{-\boldsymbol{W}_p^{\mathrm{T}} \boldsymbol{d}_p}}{e^{\boldsymbol{W}_p^{\mathrm{T}} \boldsymbol{d}_p} + e^{-\boldsymbol{W}_p^{\mathrm{T}} \boldsymbol{d}_p}}, \tag{3}$$

where $\boldsymbol{W}_p$ indicates the weight of predicted five personality scores $\hat{\boldsymbol{s}}_i^p$ from the last FC layer $\boldsymbol{d}_p$. We adopt the Euclidean loss to optimize the personality prediction task, and it is formulated as

$$L_p = \frac{1}{N_p} \sum_{i=1}^{N_p} \sum_{j=1}^{5} \|\hat{s}_{i,j}^p - s_{i,j}^p\|_2^2, \tag{4}$$

where N_p is the number of images for training and $\{s_{i,j}^p\}_{j=1}^5$ are the predicted five personality scores of the i-th image.

Inter-task Correlation Learning. In order to learn the relationship between the personality task and the aesthetic task, we introduce an inter-task correlation learning to explore the contribution of each personality trait in modeling generic image aesthetics. For the i-th image, the final aesthetic score $\hat{s}_i$ can be calculated by

$$\hat{s}_i = \hat{s}_i^a + \sum_{j=1}^{5} w_j \hat{s}_{i,j}^p, \tag{5}$$

where $\{w_j\}_{j=1}^5$ are the weights of five personality scores, $\hat{s}_i^a$ and $\{\hat{s}_{i,j}^p\}_{j=1}^5$ indicate the predicted score from aesthetic task and BF traits from personality task, respectively. Then, the Euclidean loss is employed to optimize the inter-task correlation learning model, and it is defined as

$$L = \frac{1}{N}\sum_{i=1}^{N} \|\hat{s}_i - s_i\|_2^2, \quad (6)$$

where N is the number of images for training and s_i is the average aesthetic score for the i-th image.

In the training process, the actual aesthetic scores and personality scores are normalized in the range $[0, 1]$ and $[-1, 1]$, respectively. We leverage two domain data to alternately learn both aesthetic task and personality task by optimizing loss functions (L_a or L_p). Then, we fix all the earlier layers of the two tasks, and fine-tune the inter-task correlation model with aesthetic domain data to predict the final generic aesthetic score by optimizing the loss function L. Finally, the generic image aesthetic quality can be predicted by our end-to-end multi-task model.

4. EXPERIMENTS

4.1. Experimental Setting

Implementation Details. The shared layers of the aesthetic task and personality task are from DenseNet121 [22], which is pre-trained on ImageNet [13]. We re-size the input images to 256×256 and randomly crop into 224×224 to feed into the network. In the aesthetic task and personality task, the initial learning rates of shared layers and prediction layers are set to 0.00001 and 0.0001, respectively. In the inter-task correlation network, the initial learning rate is 0.0001. The learning rate drops to a factor of 0.9 every epoch. The remaining hyper-parameters are set as follows: weight decay of $1e-5$, momentum of 0.9, and total epoch of 20. We use Pytorch [23] to implement the proposed method.

Aesthetic Databases. We evaluate the performance of our approach on two public image aesthetic databases, including AVA [8] and AADB [5]. AVA database contains more than 250,000 images, each of which is rated by about 200 individual users. The aesthetic scores range from 1 to 10. Similar to [8, 14], 230,000 images are selected for aesthetic model training and the remaining 20,000 images are used for testing. AADB database contains around 10,000 images, each of which is labeled with aesthetic score and ten aesthetic attributes by five workers. The aesthetic scores are rated from 1 to 5. Similar to [5], the database is split into three subsets, including 8,500 training images, 500 validation images, and 1000 testing images. In the training process, the average aesthetic scores are used as the supervision of images to learn the generic aesthetic model.

Personality Database. We leverage the PsychoFlickr database [12], which contains users liked images and their associated BF personality traits, to learn the personality attributes of images. In the PsychoFlickr database, 60,000 liked images of 300 Flickr individual users (200 images per user) are collected. Each individual's BF personality traits, Openness (O), Conscientiousness (C), Extroversion (E), Agreeableness (A), and Neuroticism (N), are collected by BIF-10 [18] questionnaire. All liked images with the personality traits of users who like these images are used to learn the task for personality prediction.

Table 1. Performance comparison of different methods on AVA database.

Methods	ACC(%)
AVA handcrafted features [8]	68.0
RAPID [4]	75.4
DMA [14]	75.4
Kao *et al.* [20]	76.2
Wang *et al.* [15]	76.9
Kong *et al.* [5]	77.3
Kao *et al.* [21]	79.1
Proposed (only aesthetic task)	76.1
Proposed	**81.5**

4.2. Performance Evaluation on AVA Database

Baseline Methods and Performance Criteria. To validate the performance of our method for generic image aesthetic assessment, we conduct comparative experiments with seven state-of-the-art methods [4, 5, 8, 15, 14, 20, 21]. For fair comparison with the existing classification results reported on the AVA database, we simply threshold the final estimated scores $\hat{s}_i$ to produce a binary classification. The threshold of low and high aesthetic scores is set to 0.5. For classification, the overall accuracy (ACC) is the most popular metric to evaluate the classification performance of algorithms. Higher ACC value represents better performance.

Table 1 summarizes the performance of different methods on AVA database. From this table, it can be observed that the proposed method outperforms the other seven methods. Furthermore, when only aesthetic task is used for modeling image aesthetics, our method can also achieve competitive performance compared with the state-of-the-arts. After incorporating the personality task, the overall accuracy of our method increases by 5.4% (from 76.1% to 81.5%). This demonstrates the effectiveness of our approach for assessing the generic aesthetic quality of images by taking advantage of personality traits. Apart from directly learning an independent aesthetic task, our multi-task learning approach can leverage inherent information from two domain data (aesthetics and personality) to improve the performance for generic image aesthetic assessment.

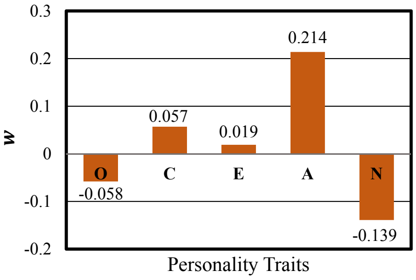

Fig. 4. The coefficients of five personality traits learned by the inter-task model.

4.3. Performance Evaluation on AADB Database

Baseline Methods and Performance Criteria. Instead of formulating image aesthetics as a binary classification issue in [4, 5, 8, 15, 14, 20, 21], our method predicts the generic image aesthetic score based on regression. For comparison with the only existing regression methods reported on the AADB database [5], we further conduct comparative experiments to validate the performance of our method with the state-of-the-art approaches, including the alone regression model (Reg), regression ranking model (Reg+Rank) and attribute-adaptive with content-adaptive model (Reg+Rank+Att+Cont) in [5]. Spearman rank order correlation coefficient (SROCC) is used to evaluate the prediction monotonicity. Pearson linear correlation coefficient (PLCC) and root mean squared error (RMSE) are used to evaluate the prediction accuracy. Higher value represents better performance for SROCC and PLCC, while lower value indicates better performance for RMSE.

In Table 2, we list comparative performance of several approaches on AADB database. We can find that the proposed method only with aesthetic task has a better performance than Reg [5]. This indicates that our basic model DenseNet [22] is more effective than AlexNet [13] used in [5] for learning the visual aesthetic representation. Our multi-task model can achieve higher SROCC values, higher PLCC values and lower RMSE values than Reg+Rank+Att+Cont [5], which takes advantage of the annotated aesthetic attributes and contents for visual aesthetic assessment. This demonstrates that the personality task in our multi-task model can provide very helpful information in modeling the generic image aesthetics.

Table 2. Performance comparison of different methods on AADB database.

Methods	SROCC	PLCC	RMSE
Reg [5]	0.624	0.618	0.029
Reg+Rank [5]	0.652	0.657	0.024
Reg+Rank+Att+Cont [5]	0.678	0.684	0.019
Proposed (only aesthetic task)	0.637	0.668	0.022
Proposed	**0.680**	**0.702**	**0.016**

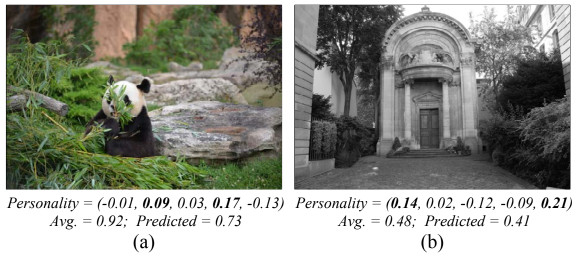

Fig. 5. Two test images in AADB database: (a) an image with high avergae aesthetic score; (b) an image with low avergae aesthetic score. The order of personality is (O, C, E, A, N), indicating the traits of users who like this image. The aesthetic scores range from 0 to 1 and the personality scores range from -1 to 1.

4.4. Discussion

To investigate how personality traits influence humans' aesthetic preferences on images, the coefficients of the five personality traits ($\{w_j\}_{j=1}^5$) learned by the inter-task model on AADB database are shown in Figure 4. The coefficient indicates the correlation between personality traits and image aesthetics. The personality traits "Conscientiousness", "Extroversion" and "Agreeableness" have positive correlation with image aesthetics, which indicates that if an image is mostly liked by people with these three personality traits, its aesthetic score tends to be rated high, especially for "Agreeableness". By contrast, The personality traits "Openness" and "Neuroticism" have negative correlation with image aesthetics, which indicates that if an image is mostly liked by people with these two personality traits, its aesthetic score tends to be rated low, particularly for "Neuroticism". Figure 5 shows two test images with the predicted personality and aesthetic score in AADB database. As shown in Figure 5(a), the image with high aesthetics is liked by users with high "Agreeableness" and "Conscientiousness". The image with low aesthetics is shown Figure 5(b), which is liked by users with high "Openness" and "Neuroticism". It also verifies that the personality attributes of images can capture aesthetic differences in learning the generic aesthetic model. In addition, the predicted aesthetic scores of our multi-task model have a high consistency with the average aesthetic scores.

5. CONCLUSIONS

In this paper, we have proposed an end-to-end personality driven multi-task deep learning model for generic image aesthetic assessment. A multi-task CNN model is developed to address aesthetic task and personality task simultaneously. Based on the relationship between personality traits and image aesthetics, an inter-task correlation learning is introduced to our multi-task model for image aesthetic assessment.

The personality traits have been shown effective in capturing individuals' aesthetic differences on images, which are discriminative representations to optimize visual aesthetics model for the "average" user. Experimental results on two public databases have demonstrated that our approach is superior to the state-of-the-arts.

6. REFERENCES

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

[2] W. T. Sun, T. H. Chao, Y. H. Kuo, and W. H. Hsu, "Photo filter recommendation by category-aware aesthetic learning," *IEEE Transactions on Multimedia*, vol. 19, no. 8, pp. 1870–1880, 2016.

[3] W. Wang, J. Shen, and H. Ling, "A deep network solution for attention and aesthetics aware photo cropping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, doi:10.1109/TPAMI.2018.2840724.

[4] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rating pictorial aesthetics using deep learning," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 457–466, 2014.

[5] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *European Conference on Computer Vision*, 2016, pp. 662–679.

[6] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *European Conference on Computer Vision*, 2006, pp. 288–301.

[7] Xiaoou Tang, Wei Luo, and Xiaogang Wang, "Content-based photo quality assessment," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 1930–1943, 2013.

[8] N. Murray, L. Marchesotti, and Perronnin F., "Ava: A large-scale database for aesthetic visual analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2408–2415.

[9] 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 Transactions on Multimedia Computing, Communications, and Applications*, vol. 14, no. 2s, pp. 30:1–30:24, 2018.

[10] H. Zhu, L. Li, and H. Jiang, "Evaluating attributed personality traits from scene perception probability," *Pattern Recognition Letters*, vol. 116, pp. 121–126, 2018.

[11] 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," in *Neural Processing Letters*, 2019, doi:10.1007/s11063-019-09987-7.

[12] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina, "Unveiling the multimedia unconscious," in *ACM International Conference on Multimedia*, 2013, pp. 213–222.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[14] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *IEEE International Conference on Computer Vision*, 2015, pp. 990–998.

[15] W. Wang, M. Zhao, L. Wang, J. Huang, C. Cai, and X. Xu, "A multi-scene deep learning model for image aesthetic evaluation," *Signal Processing Image Communication*, vol. 47, no. C, pp. 511–518, 2016.

[16] G. Matthews and I. J Deary, "Personality traits.," *Cambridge University Press*, vol. 28, no. 27, pp. 325, 2009.

[17] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

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

[19] R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997.

[20] Y. Kao, K. Huang, and S. Maybank, "Hierarchical aesthetic quality assessment using deep convolutional neural networks," *Signal Processing Image Communication*, vol. 47, pp. 500–510, 2016.

[21] Y Kao, R. He, and K. Huang, "Deep aesthetic quality assessment with semantic information," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1482–1495, 2017.

[22] Huang G., Liu Z., Maaten L. V. D., and Weinberger K. Q., "Densely connected convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2261–2269.

[23] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *International Conference on Neural Information Processing Systems Workshop*, 2017.