*Erschienen in: ATQAM/MAST'20: Joint Workshop on Aesthetic and Technical Quality Assessment of Multimedia and Media Analytics for Societal Trends / Hosu, Vlad et al. (Hrsg.). - New York : ACM, 2020. - S. 19-20. - ISBN 978-1-4503-8154-3*
*https://dx.doi.org/10.1145/3423268.3423589*

# From Technical to Aesthetics Quality Assessment and Beyond: Challenges and Potential

Vlad Hosu
Dietmar Saupe
Bastian Goldlücke
University of Konstanz
Konstanz, Germany

Weisi Lin
Nanyang Technological University
Singapore

Wen-Huang Cheng
National Chiao Tung University
Hsinchu, Taiwan

John See
Lai-Kuan Wong
Multimedia University
Cyberjaya, Malaysia

## ABSTRACT

Every day 1.8+ billion images are being uploaded to Facebook, Instagram, Flickr, Snapchat, and WhatsApp [6]. The exponential growth of visual media has made quality assessment become increasingly important for various applications, from image acquisition, synthesis, restoration, and enhancement, to image search and retrieval, storage, and recognition.

There have been two related but different classes of visual quality assessment techniques: image quality assessment (IQA) and image aesthetics assessment (IAA). As perceptual assessment tasks, subjective IQA and IAA share some common underlying factors that affect user judgments. Moreover, they are similar in methodology (especially NR-IQA in-the-wild and IAA). However, the emphasis for each is different: IQA focuses on low-level defects e.g. processing artefacts, noise, and blur, while IAA puts more emphasis on abstract and higher-level concepts that capture the subjective aesthetics experience, e.g. established photographic rules encompassing lighting, composition, and colors, and personalized factors such as personality, cultural background, age, and emotion.

IQA has been studied extensively over the last decades [3, 14, 22]. There are three main types of IQA methods: full-reference (FR), reduced-reference (RR), and no-reference (NR). Among these, NR-IQA is the most challenging as it does not depend on reference images or impose strict assumptions on the distortion types and level. NR-IQA techniques can be further divided into those that predict the global image score [1, 2, 10, 17, 26] and patch-based IQA [23, 25], naming a few of the more recent approaches.

Comparatively, IAA received less research attention than IQA until recently [5]. Pioneering IAA approaches adopted handcrafted features for training an aesthetics model, while earlier approaches used global and regional features [4, 12] and the later approaches employed subject-focused features [16, 24] and generic descriptors [18]. With the emergence of the deep learning techniques, Lu et al. [15] proposed RAPID, the first DCNN-based model for aesthetics. Subsequently, many other DCNN-based aesthetics models with an improved performance have been proposed [7, 8, 21]. More recently, motivated by the subjectivity of aesthetic preference among individuals, [13, 20] attempted to model personalized image aesthetics.

Even though IQA and IAA have mostly been studied independently, they represent tightly related aspects of the same underlying subjective experience of media items: value judgments. Several works have attempted to bridge the gap between these two fields. Judith & Ingrid [19] studied to what extent the presence of artifacts impacts the aesthetic quality of images. Their experiments show that image integrity somewhat influences aesthetic quality scores. Jenadeleh et al. [11] believe that visual quality highly correlates with the human judgment of the image quality and encompasses perceptual image qualities such as aesthetics, semantics, context, and various types of visual distortion. They proposed a schema that uses a set of aesthetics features to improve NR-IQA performance, which is confirmed by experimental results. [8, 21] proposed DCNN aesthetics models that are cross-tested on IQA datasets. Notably, all these works only loosely linked IQA and IAA.

The pipeline for building an algorithmic assessment model is the same for both IQA and IAA. It starts with performing subjective studies to annotate image datasets with ratings. The latter is used to train machine learning models with objectives helpful for various applications. This pipeline and its components present individual and overarching challenges for both fields. We believe that understanding the relation between IQA and IAA and conducting a joint study of these two fields can lead to opportunities for further research and accelerate the progress of research in both fields.

**Challenges:** Ideally, subjective studies should be performed in such a way that the labels collected are unbiased and the annotation is reliable. Machine learning (ML) models, especially deep learning approaches perform best when trained on large and sufficiently diverse datasets. The ML methods themselves should generalize well, from a few examples of each degradation/enhancement type, rating range, etc. However, each of these components has its limitations.

First, the task definition used in subjective studies is often ambiguous and subject to interpretation leading to biases in the opinions expressed by participants. On a similar note, the best option

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.



*Konstanzer Online-Publikations-System (KOPS)*
*URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-3gz5dg09sllp3*

for creating large datasets is crowdsourcing; however, the study methodology often results in unreliable annotations due to the relationship between control and bias: too much or too little control introduces different kinds of biases. Second, machine learning approaches such as deep learning tend to overfit to the data domain, resulting in cross-database performance loss and a reduced generalization in-the-wild; this is partially due to the methods themselves, e.g. adversarial examples, but also due to the domain or label-space shifts. To conclude, subjective labels are inaccurate to start with, data annotation is expensive, and the factors compound with deficiencies of ML methods leading to sub-optimal assessment models.

**Potential Opportunities:** Studying the relationship between IQA and IAA can lead to novel solutions for the previously mentioned challenges. Moreover, proven approaches that work well for IQA can be applied/transferred to IAA.

A better understanding of the common factors for IAA/IQA can lead to improved task definitions in subjective studies by separating the technical and aesthetic aspects. Understanding relationships between IAA and IQA can be done by creating suitable datasets for both tasks, that ensure the diversity of ratings in both the technical and aesthetics domain, e.g. all combinations, such as low/high technical + low/high aesthetics.

Artificial degradations have been used as objective ground-truth for controlling subjective studies [9], and have been shown to improve reliability. Analogous artificial degradations could be useful to define control questions for aesthetics as well. This approach can also provide new ways to generate datasets: similar to artificial degradations in IQA, synthetic aesthetics enhancements/degradations can be introduced, e.g. crops, rules of thirds, etc.

Furthermore, assessment can focus on new objectives such as joint prediction of IAA and IQA, personalized prediction of ratings or fine-grained aesthetics relationships.

## CCS CONCEPTS

• **Computing methodologies** → *Perception*; *Image manipulation*; **Computer vision tasks**; • **Information systems** → **Multimedia information systems**; *Data mining*.

## KEYWORDS

image quality assessment, image aesthetics assessment, IQA, IAA, potential, challenges

## ACKNOWLEDGMENT

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 251654672, TRR 161 (Project A05).

## REFERENCES

[1] Shahrukh Athar and Zhou Wang. 2019. A Comprehensive Performance Evaluation of Image Quality Assessment Algorithms. *Ieee Access* 7 (2019), 140030–140070.

[2] Simone Bianco, Luigi Celona, Paolo Napoletano, and Raimondo Schettini. 2018. On the use of deep learning for blind image quality assessment. *Signal, Image and Video Processing* 12, 2 (2018), 355–362.

[3] Damon M Chandler. 2013. Seven challenges in image quality assessment: past, present, and future research. *International Scholarly Research Notices* 2013 (2013).

[4] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. 2006. Studying aesthetics in photographic images using a computational approach. In *European conference on computer vision*. Springer, 288–301.

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

[6] Jim Edwards. 2014 (accessed September 15, 2020). *PLANET SELFIE: We're Now Posting A Staggering 1.8 Billion Photos Every Day.* https://www.businessinsider.com/were-now-posting-a-staggering-18-billion-photos-to-social-media-every-day-2014-5

[7] Yong-Lian Hii, John See, Magzhan Kairanbay, and Lai-Kuan Wong. 2017. Multigap: Multi-pooled inception network with text augmentation for aesthetic prediction of photographs. In *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1722–1726.

[8] Vlad Hosu, Bastian Goldlucke, and Dietmar Saupe. 2019. Effective aesthetics prediction with multi-level spatially pooled features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 9375–9383.

[9] Vlad Hosu, Hanhe Lin, and Dietmar Saupe. 2018. Expertise screening in crowdsourcing image quality. In *QoMEX 2018: Tenth International Conference on Quality of Multimedia Experience*.

[10] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe. 2020. KonIQ-10k: An Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment. *IEEE Transactions on Image Processing* 29 (2020), 4041–4056.

[11] Mohsen Jenadeleh, Mohammad M Masaeli, and Mohsen E Moghaddam. 2017. Blind image quality assessment based on aesthetic and statistical quality-aware features. *Journal of Electronic Imaging* 26, 4 (2017), 043018.

[12] Yan Ke, Xiaoou Tang, and Feng Jing. 2006. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 1. IEEE, 419–426.

[13] 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.

[14] Weisi Lin and C-C Jay Kuo. 2011. Perceptual visual quality metrics: A survey. *Journal of visual communication and image representation* 22, 4 (2011), 297–312.

[15] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James Z Wang. 2014. Rapid: Rating pictorial aesthetics using deep learning. In *Proceedings of the 22nd ACM international conference on Multimedia*. 457–466.

[16] Yiwen Luo and Xiaoou Tang. 2008. Photo and video quality evaluation: Focusing on the subject. In *European Conference on Computer Vision*. Springer, 386–399.

[17] Kede Ma, Wentao Liu, Tongliang Liu, Zhou Wang, and Dacheng Tao. 2017. dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Transactions on Image Processing* 26, 8 (2017), 3951–3964.

[18] Luca Marchesotti, Florent Perronnin, Diane Larlus, and Gabriela Csurka. 2011. Assessing the aesthetic quality of photographs using generic image descriptors. In *2011 international conference on computer vision*. IEEE, 1784–1791.

[19] Judith A Redi and Ingrid Heynderickx. 2012. Image integrity and aesthetics: towards a more encompassing definition of visual quality. In *Human Vision and Electronic Imaging XVII*, Vol. 8291. International Society for Optics and Photonics, 829115.

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

[21] Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural image assessment. *IEEE Transactions on Image Processing* 27, 8 (2018), 3998–4011.

[22] Zhou Wang and Alan C Bovik. 2006. Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing* 2, 1 (2006), 1–156.

[23] Oliver Wiedemann, Vlad Hosu, Hanhe Lin, and Dietmar Saupe. 2018. Disregarding the Big Picture: Towards Local Image Quality Assessment. In *10th International Conference on Quality of Multimedia Experience(QoMEX)*. IEEE. http://database.mmsp-kn.de

[24] Lai-Kuan Wong and Kok-Lim Low. 2009. Saliency-enhanced image aesthetics class prediction. In *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 997–1000.

[25] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. 2020. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3575–3585.

[26] Weixia Zhang, Kede Ma, and Xiaokang Yang. 2019. Learning to Blindly Assess Image Quality in the Laboratory and Wild. *arXiv preprint arXiv:1907.00516* (2019).