Can We Measure the Beauty of an Image?

Matteo Bodini*

Department of Computer Science "Giovanni Degli Antoni", University of Milan, Italy

*Corresponding author: Matteo Bodini, Department of Computer Science "Giovanni Degli Antoni", University of Milan, Italy.

Abstract

Although the concept of image quality has been a subject of study for the image processing community for more than forty years, notions related to aesthetics of pictures have only appeared in the last ten years. Studies devoted to this theme are growing today, taking advantage of machine learning techniques, and the proliferation of specialized websites for archiving photos on the Internet. Before examining the progress of such computer methods, we recall the earlier approaches to aesthetics measurement. Then, we take a look at the work done in the community of pattern recognition and artificial intelligence, we compare the presented results, and we critically examine the steps taken.

Introduction

In the last three years, several communications have proliferated in the pattern recognition and artificial intelligence community with the aim to provide an automatic assessment of the aesthetic value of an image [1-7]. These works make use of deep neural networks (DNNs), and they follow a handful of works, dated at the beginning of the century, which tackled this problem with the help of more classical techniques of learning: detection of primitives chosen by the user (handcrafted), and classifiers of various types [8-13]. As always, DNNs techniques have quickly outperformed the more traditional methods, as they have done in many other areas of pattern recognition. However, this has only been made possible by a number of studies, taken place in a scientific context in which the problem can be approached in many different ways, in particular by knowledge of neurobiology, as well as by experiments of social psychology.

The first works that proposed a mathematical measure of beauty are due to Charles Henry [14] but is the mathematician George D. Birkhoff who proposed the first operational formulation [15]. This formulation, inspired by 25 centuries of philosophical literature on aesthetics, in particular in visual arts, was built on notions of order and simplicity at a time when these two terms had little meaning in mathematics. The idea of Birkhoff has been enriched over the last century by successive contributions of Gestalt psychology, information theory, mathematical morphology, and the theory of

complexity to arrive at algorithmic and algebraic expressions [16-18], which had very interesting results.

Techniques based on machine learning, that began with this century, have made a clean sweep of this work. They are opportunely exploiting new resources: the abundance of images accessible on the Internet, the availability of many sources of expertise through specialized social networks or the well-known general public ones, the appearance of powerful statistical techniques for learning rules of classifications, and finally the possibility of extending them to large unknown groups. Further, the diffusion of DNNs, successively exploiting convolutional filtering and fully connected layers, has often limited the human expertise only to the constitution of big indexed datasets, fundamental in learning.

This complete break in the paradigms behind the aesthetic approach should be analyzed and should be considered the consequences that can be expected from it. It is also fundamental to compare the expectations of the users with the potential results of the employed methods.

Machine Learning Approaches

As we have said, the first aesthetic measurement works were purely algebraic (they did not use machine learning techniques) and did not result in a very large com-munity agreement. The following machine learning based works started on very different bases, and they are presented in Table 1.

Table 1: System for learning aesthetics, their learning datasets, and methods used for classification.

Ref.	Dataset	Primitives and classifier
[8]	DP Challenge	generic attributes – Bayes, AdaBoost
[19]	DP Challenge	image primitives – Bayes, SMV, AdaBoost
[9]	Photo.net	image primitives – SVM
[10]	CUHK	generic attributes – PCA + FV + SVM
[20]	DP Challenge	high level primitives – SVM
[21]	Photo.net	texture, color – Acquine + SIMPLIcity
[22]	CUHK	image primitives – SVM
[23]	DP Challenge	internet users' opinions – SV-epsilon
[24]	AVA	HSV or Munsell – LDA + Lasso
[1]	AVA	image + ss-image – DNN/SCNN + RD-CNN
[25]	BEAUTY	image primitives – PLSR
[26]	AADB + AVA	reduced image + categories – DNN
[3]	AVA	pyramid + class information – multitask MNA DNN
[27]	AVA	image + 3 categories – A & C DNN
[4]	AVA	histogram of ratings – brain-inspired DNN
[28]	Redi	web images – DNN
[5]	AVA	categories – multitask DNN
[6]	AVA	composition + ss-images – multitask A-Lamp
[7]	AVA	histogram of ratings + categories – multitask DNN

Handcrafted features and classification

The first publications aimed at automatically learning aesthetics rules in images are dated back to 2006 [29,8]. These works are based on the availability of numerous datasets of images on the Internet, which that are often provided with information for judging the beauty of each image. The nature of these datasets is very variable. Some are simple archives exchanged between individuals (Flickr, for instance), others are intended for amateurs or even professionals, and photographers (such as Photo.net or DP Challenge), while others are reserved to scientists (Image CLEF, Beauty, AVA). The annotations reflect very diverse opinions, from the like of the general public, to expert opinions of photographic juries, by specialized consultations. The several works presented in this context differ in the choice of primitives, and in the choice of the classifier [21,30,25,31]. The primitives conventionally used in the field of multimedia [9,32-36] have been used concurrently with primitives more specialized in the field of aesthetics [19,20,24], without showing any loss of performance. The classifiers used were based on Bayesian techniques, SVM, or decision graphs.

Deep neural networks

From their appearance in the field of aesthetic evaluation, DNNs-based techniques have shown superior performance over more conventional machine learning approaches. The architectures adopted are those found throughout the field of recognition in images: layers of convolutions, followed by totally connected layers, or, more recently, only convolutional layers. However, some changes have been proposed to adapt these systems to the specificities of the problem: several works have been suggested to treat very large

images while preserving the fine structure of the details: window pre-selection around points of interest [6,37], parallel processing of randomly drawn windows [24], use of hierarchical structures [38,3], etc. Despite the latter, the size of the operational DNNs input layers is a limit for works on aesthetics that still handle large images; taking into account additional information, which is very important in the choice of criteria to be applied, has led to multiple flow networks [5,6] which exploit various knowledge: the type of image, the style of the photo, the class of the main object, etc.; the reproduction of certain brain mechanisms has led to the separation of the processing architecture in different ways [21,5] or, sometimes, in a succession of DNNs: one in charge of the low level, and another in charge of the information of the high level [4].

The implementation of DNNs-based techniques has significantly changed the work done on the aesthetics of images. A first element of differentiation concerns the choice of datasets. The need for very large learning datasets has led to the abandonment previously used original datasets, composed only by a few thousand images. The community has thus focused on the AVA dataset which has the advantage of having images often very beautiful, with many opinions on each one. However, for training networks, its size (it is made up of 250 000 images) is often insufficient. One can then resort to artificial extension by manipulating images (data augmentation) [4]. A second element is the almost complete disappearance (except in [6]) of the aesthetic criteria for the construction of the DNNs architecture. Works that rely on information external to the image mainly use data coming from the type of image: interior, portrait, sport, etc., data that seem, however, quite distantly related to the beauty of the image.

Analysis of the Works

The work carried out by the image processing community can be exploited to evaluate the beauty of photos from several point of view. Here we select and analyze the main interesting points of these works.

Evaluation of performances

Aesthetic studies are very difficult to evaluate. The database (AVA [11]) was created to alleviate this problem. It has been endowed with a very large body of quality annotations on a scale of 1 to 10 (almost 200 notes per image). This dataset is often used by matching the samples in two families: the beautiful images (whose average score is greater than $5 + d$) and the ugly images (whose average score is lower than $5\ d$). The parameter d allows each one to adorn the classes during the learning with the rigor that he wishes. It is often taken as zero or equal to 1. If we consider $d=0$, we can notice a steady growth trend in recognition performances.

Continuous image classification

From the beginning [29], many works had the aim of classifying images according to a scale of beauty more or less continuous. Although many algorithms provide a score between 0 and 10, few studies report the quality of these notations [7], except to refine the binary decision [27,4]. Evaluation of a continuous ranking is very

difficult today and seems to us a major issue. Let us note that in [25] a classification in five levels makes it possible to substantially refine the measurement. Note especially the very original approach of [39], which proposes to compare pairwise the images of the dataset to reach a relative evaluation.

Properties that should be further exploited

The analysis of an image by the classical architectures of DNNs showed their power to recognize and locate objects, even deformed or partially occluded. However, it seems that some important properties of the aesthetic evaluation would require to evolve such architectures. We have already pointed out the importance of being able to process large images with many fine details. Note also the importance that should be given to chromatic harmony, which is undeniably an important component of aesthetics (the work of [37] is exemplary). It is not obvious that architectures that carry out convolutions in the first layers respect the subtlety of the nuances. The internal construction of an image is itself an important element of the aesthetic quality of the. Let us recognize that, although many works try to take account of it, very few give themselves the means to do it through the initially convoluted layers then totally connected of the DNN. To our knowledge only the authors of [6] reserve a way of treatment to this structure.

Beautiful vs ugly images

The binary criterion is adopted by the community to compare the various approaches. It can be applied quickly on very large datasets; it moves easily from one dataset to another; it lends itself to a simple visual check; it offers a good solution to some of the problems that society poses: sort very quickly large archives to keep a quintessence, provide attractive examples for illustrations, assist an operator for his shooting, etc.

It suffers, however, from being too much simplified. It is based on the assumption that all images come from one or the other category, a postulate of which no trace is found in the literature. Moreover, it is very commonly accepted, both in philosophy and neurobiology, that the attribute of beauty has only a positive valence and no equivalent to a negative valence (which would be called the ugly), being supported by other attributes like "scary", "sad", "boring", "banal", "rough", etc. Thus, the complexity of the information transmitted through the notations for each image of the AVA database is today insufficiently taken into account, even if some works try to exploit them [40,7,4].

Which beauty, and what expert?

The images used for performance tests represent what we can expect from quality images from social networks. Beautiful pictures are undoubtedly generally superior to ugly pictures. If the qualities of beautiful images are not always obvious, we see that they rarely show the flaws that make ugly images stand out: poor composition, poor chromatic distribution, lack of focus, etc. An attentive and demanding observer will, however, often disagree with the decisions made by the system, even if these decisions are in accordance with the judgments of the experts. This is often

explained either because a beautiful image is commonplace or, especially, because a quality image has been classified ugly. In the latter case it is frequently observed that original aspects of the image have been ignored. DNNs favor "normal" images, which is hardly in line with the experts' recommendations.

We regret, in this respect, that no system has confronted the reputedly remarkable images of the photographic archives. There would certainly be much to learn from the objective analysis of such results. Now, one of the most sensitive points of the DNN approach is addressed. The importance of having a dataset of high quality has been less considered in the implementation of the handcrafted features approaches, but it has become crucial for DNNs approaches. The AVA database [11] provided a good answer to this request. Beyond the large collection of images, AVA has several information attached to each photo: the evaluations, the theme covered by the image (among more than 900, from the competitions of the photo bank DP Challenge of origin), a semantic annotation (among 66, issue of the themes) and the photographic style (ascribed by professional photographers, among 14).

Is it sufficient? It is not certain. For sure for the objectivists, who place all the beauty in the only object, there is in AVA the object faithfully reproduced and, in the average, the expression of the consensus on its appreciation. There are therefore all the elements sufficient to allow a machine to reproduce the human judgment. If is given a more important place to the observer, the focus is more on the information that will be needed for the evaluation. What we can draw of this information from the AVA dataset is very doubtful. Thus, in [26], it was considered necessary to build a dataset (AADB), different from AVA, keeping the evaluator's mark during the evaluation, so that the evaluation machine could be trained. The authors indicate that such choice makes possible to obtain a better satisfaction between the rankings obtained by the same expert. In [25] the authors focused mainly on the cultural context of the experts, to build the BEAUTY dataset. Only users from a small number of countries with a high degree of cultural homogeneity were selected, and their opinions were subsequently screened to discard the deviant points of view.

Conclusion

The success of methods for evaluating the beauty of images is certain. Performed on a very large number of images, they allow to separate with reasonable performances the most beautiful images by the ugly ones. There is no doubt that these performances will improve over time, as the works currently being presented still have many margins of progress. However, let us say that today these methods are most used to elaborate a first sorting on large quantities of images. If you really want to distinguish the most beautiful images, it is still necessary to return to such sort to select the small number that surpasses the others. We regret, as we do for many other pattern recognition problems, that DNNs-based solutions are delivered to us without explicit intermediate decision steps. Thus, if we know how to sort the images, we do not really know how this sorting is done. It is for our understanding a step back from first approaches based on handcrafted primitives. Finally,

let us remark the fact that the implemented methods to date have totally ignored an important part of the aesthetic judgment that the literature puts forward: the cultural and socio-educational context of the observer [41]. This forgetfulness is understandable because, if aesthetics is a complex and poorly understood field, culture is still much more complex and poorly modeled.

Acknowledgement

Conflict of Interest

No conflict of interest.

References

1. Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, James Z Wang (2014) Rapid: Rating pictorial aesthetics using deep learning. In Proceedings of the 22nd ACM international conference on Multimedia, ACM, pp. 457-466.

2. Xin Lu, Zhe Lin, Xiaohui Shen, Radomir Mech, James Z Wang (2015) Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In Proceedings of the IEEE International Conference on Computer Vision, pp. 990-998.

3. Long Mai, Hailin Jin, and Feng Liu (2016) Composition-preserving deep photo aesthetics assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 497-506.

4. Zhangyang Wang, Shiyu Chang, Florin Dolcos, Diane Beck, Ding Liu (2016) Brain-inspired deep networks for image aesthetics assessment. ar Xiv preprint arXiv: 1601.04155.

5. Yueying Kao, Ran He, Kaiqi Huang (2017) Deep aesthetic quality assessment with semantic information. IEEE Transactions on Image Processing 26(3): 1482-1495.

6. Shuang Ma, Jing Liu, Chang Wen Chen (2017) A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: 4535-4544.

7. Hossein Talebi, Peyman Milanfar (2018) Nima: Neural image assessment. IEEE Transactions on Image Processing 27(8): 3998-4011.

8. Yan Ke, Xiaoou Tang, Feng Jing (2006) The design of high-level features for photo quality assessment. In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on 1: pp. 419-426.

9. Ritendra Datta, James Z Wang (2010) Acquine: aesthetic quality inference engine-real-time automatic rating of photo aesthetics. In Proceedings of the international conference on Multimedia information retrieval, ACM, pp. 421-424.

10. Luca Marchesotti, Florent Perronnin, Diane Larlus, Gabriela Csurka (2011) Assessing the aesthetic quality of photographs using generic image descriptors. In Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, pp. 1784-1791.

11. Naila Murray, Luca Marchesotti, Florent Perronnin (2012) Ava: A large-scale database for aesthetic visual analysis. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEE, pp. 2408–2415.

12. Matteo Bodini, Alessandro D'Amelio, Giuliano Grossi, Raffaella Lanzarotti, Jianyi Lin (2018) Single sample face recognition by sparse recovery of deep-learned lda features. In: Jacques Matteo Bodini, Blanc-Talon, David Helbert, Wilfried Philips, Dan Popescu, Paul Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems, Cham, Springer International Publishing, pp. 297-308.

13. Matteo Bodini (2019) A review of facial landmark extraction in 2d images and videos using deep learning. Big Data and Cognitive Computing 3(1): 14.

14. Charles Henry (1885) Introduction a` une esthetique´ scientifique. La revue contemporaine.

15. George Birkhoff (1933) Aesthetic Measure. Harvard University Press, USA.

16. Abraham A Moles (1957) Theorie´ de l'information et perception esthetique´. Revue Philosophique de la France et de l'Etranger 147: 233-242.

17. Max Bense (1969) Einfuhrung¨ in die informationstheoretische asthetik¨ grundlegung und anwendung in der texttheorie.

18. Jaume Rigau, Miquel Feixas, Mateu Sbert (2008) Informational aesthetics measures. IEEE Computer Graphics and Applications 28(2): 2008.

19. Yiwen Luo, Xiaoou Tang (2008) Photo and video quality evaluation: Focusing on the subject. In European Conference on Computer Vision, Springer, pp. 386-399.

20. Sagnik Dhar, Vicente Ordonez, Tamara L Berg (2011) High level describable attributes for predicting aesthetics and interestingness. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, pp. 1657-1664.

21. Lei Yao, Poonam Suryanarayan, Mu Qiao, James Z Wang, Jia Li (2012) Oscar: On-site composition and aesthetics feedback through exemplars for photographers. International Journal of Computer Vision 96(3): 353-383.

22. Kuo-Yen Lo, Keng-Hao Liu, Chu-Song Chen (2012) Assessment of photo aesthetics with efficiency. In Pattern Recognition (ICPR), 2012, 21st International Conference on, IEEE, pp. 2186-2189.

23. Jose San Pedro, Tom Yeh, and Nuria Oliver (2012) Leveraging user comments for aesthetic aware image search reranking. In Proceedings of the 21st international conference on World Wide Web, ACM, pp. 439-448.

24. Peng Lu, Zhijie Kuang, Xujun Peng, Ruifan Li (2014) Discovering harmony: A hierarchical color harmony model for aesthetics assessment. In Asian Conference on Computer Vision, Springer, pp. 452-467.

25. Rossano Schifanella, Miriam Redi, Luca Maria Aiello (2015) An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures. In ICWSM, pp. 397-406.

26. Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, Charless Fowlkes (2016) Photo aesthetics ranking network with attributes and content adaptation. In European Conference on Computer Vision, Springer, pp. 662-679.

27. Yueying Kao, Kaiqi Huang, Steve Maybank (2016) Hierarchical aesthetic quality assessment using deep convolutional neural networks. Signal Processing: Image Communication 47: 500- 510.

28. Miriam Redi, Frank Z Liu, Neil O'Hare (2017) Bridging the aesthetic gap: The wild beauty of web imagery. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, ACM, pp. 242-250.

29. Ritendra Datta, Dhiraj Joshi, Jia Li, James Z Wang (2006) Studying aesthetics in photographic images using a computational approach. In European Conference on Computer Vision, pages, Springer, pp. 288-301.

30. Kuo-Yen Lo, Keng-Hao Liu, Chu-Song Chen (2012) Intelligent photographing interface with on-device aesthetic quality assessment. In Asian Conference on Computer Vision, Springer, pp. 533-544.

31. Peng Lu, Xujun Peng, Xinshan Zhu, Ruifan Li (2016) An el-lda based general color harmony model for photo aesthetics assessment. Signal Processing 120: 731-745.

32. Luca Marchesotti, Florent Perronnin, France Meylan (2013) Learning beautiful (and ugly) at-tributes. BMVC 7: 1-11.

33. Matteo Bodini (2019) Probabilistic nonlinear dimensionality reduction through gaussian process latent variable models: an overview. In First Annual Conference on Computer-aided Developments in Electronics and Communication (CADEC-2019).

34. Matteo Bodini (2019) Sound classification and localization in service robots with attention mechanisms. In First Annual Conference on Computer-aided Developments in Electronics and Communication (CADEC-2019).

35. Giuseppe Boccignone, Matteo Bodini, Vittorio Cuculo, and Giuliano Grossi (2018) Predictive sampling of facial expression dynamics driven

by a latent action space. In 2018 14th International Conference on Signal-Image Technology Internet-Based Systems (SITIS): 143-150.

36. Alessandro Rizzi, Daniela Fogli, Barbara Rita Barricelli (2017) A new approach to perceptual assessment of human-computer interfaces. Multimedia Tools and Applications 76(5): 7381-7399.

37. Peng Lu, Xujun Peng, Ruifan Li, Xiaojie Wang (2015) Towards aesthetics of image: A Bayesian framework for color harmony modeling. Signal Processing: Image Communication 39: 487-498.

38. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition.

IEEE transactions on pattern analysis and machine intelligence 37(9): 1904-1916.

39. Katharina Schwarz, Patrick Wieschollek, Hendrik PA Lensch (2016) Will people like your image? arXiv preprint arXiv: 1611.05203.

40. Bin Jin, Maria V Ortiz Segovia, and Sabine Susstrunk (2016) Image aesthetic predictors based on weighted cnns. In Image Processing (ICIP), 2016 IEEE International Conference on, Ieee, pp. 2291-2295.

41. Mathias Waschek (2000) Le chef-d'œuvre: un fait culturel. AA. VV, Qu'Est-ce qu'un Chefe-d' Œuvre.