Experimental aesthetics without semantics

C.Alejandro Parraga[1], Marcos Muñoz González[2], Xavier Otazu[1], Olivier Penacchio[3]

(1) Computer Vision Centre /Computer Science Dept., Univ. Autònoma de Barcelona, Spain;
(2) Computer Science Dept., Univ. Autònoma de Barcelona, Spain;
(3) School of Psychology and Neuroscience, University of St Andrews, UK.
http://www.cvc.uab.cat/neurobit

UAB Universitat Autònoma de Barcelona
CVC Centre de Visió per Computador
neurobit NeuroComputational and Biological Vision Team

The problem with databases…

1. Too much "semantics"

Observers can easily assess how beautiful/ugly they consider an image to be. However, this aesthetics decision is influenced by both perceptual factors (which determine bottom-up neural processing) and cognitive factors such as contextual variables (which determine personal aesthetic preferences). We call the later "image semantics". For example, a cute pet or baby might be disproportionally highly rated just because of the emotions involved.

This dualism (see Fig.1) makes the problem of predicting aesthetic responses particularly difficult from a computational point of view.

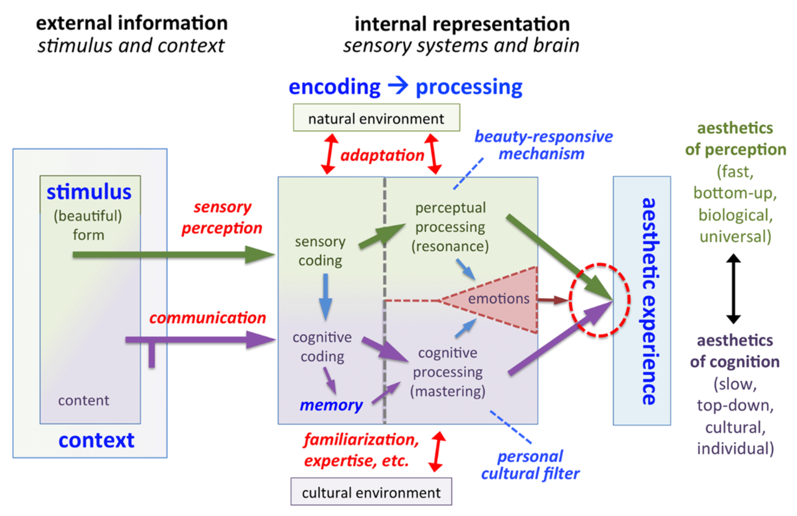
2. "Shocking" instead of "ugly"

Emotions affect the lower end of the beauty spectrum as well. A simple Google search for "ugly images" will bring us semantically- and emotionally- loaded pictures such as deformed human-like faces or a suffering animals.

3. Too little "true ugliness"?

Current image datasets (used in machine learning) are derived from online "beauty contests" o websites where amateur photographers upload their best pictures for their peers or the public to rate them. No one uploads an image he/she considers "ugly", therefore these are under-represented, which biases almost all aesthetics datasets

Figure 1



Redies C (2015) Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. Front. Hum. Neurosci. 9:218. doi: 10.3389/fnhum.2015.00218

Figure 2

Our solution: an image dataset devoid of "semantics"

1. Remove "semantics"

Our solution to the first problem above was to create a database of images devoid of contextual information (i.e., "semantics") which contains 5684 **images of natural objects.**

2. Remove "emotions"

We tackle the problem of the emotional content of our dataset by **not including images of people, animals, and other emotionally-charged motives** (such as well-known holiday scenes)

3. Add "ugly" scenes

To address the strong bias towards highly valued (beautiful) scenes, we incorporate 1791 images that were **modified by 40 volunteers using a custom-made image processing program.** They had instructions to either make the images as ugly as possible ("uglify") or to make them more beautiful ("beautify"). We incorporated 872 "uglified" images and 919 "beautified" images. To compensate for possible systematic biases, we also added 2951 randomly modified images.

In summary, we created a dataset that stimulates mainly the sensory/perceptual pathway, as described in Fig. 1

Crowdsourcing

We then ask observers to evaluate the aesthetic value of every image using a crowdsourcing paradigm (**10426 images, 100 valuations each**). Fig.3 shows the distribution of valuations.

The responses' histograms were then fitted by **Truncated-Gaussians** to obtain a single mean aesthetic valuation and its StDev per image (see Fig.4)

Figure 3

Distribution of aesthetic valuation in the dataset

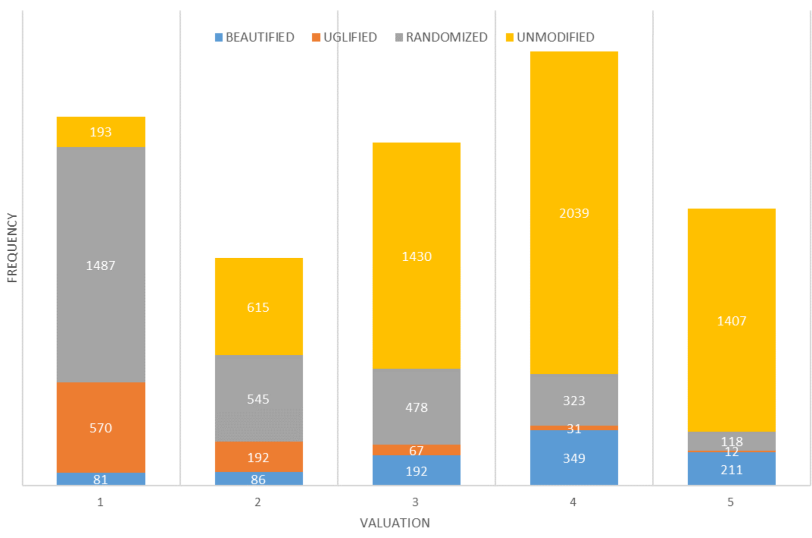
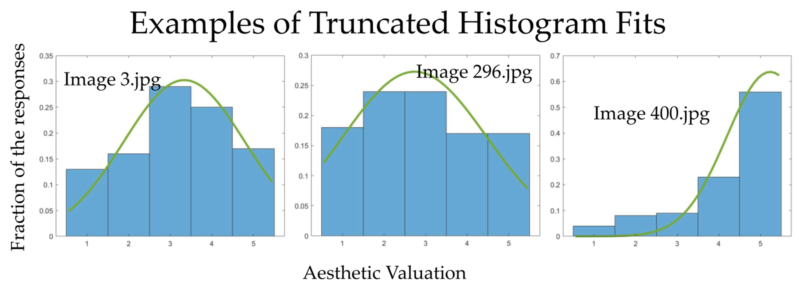


Figure 4

Examples of Truncated Histogram Fits



Results

This unbiased, low-semantics database allows us to study the low-level visual properties that are more likely to explain observers' aesthetic response.

We tried several image metrics and explored their correlation with the aesthetic valuations obtained from crowdsourcing. The most interesting ones are:

1) Chrom + Achrom contrast: calculated by applying multiscale "Mexican hat" convolutions to each of the chromatically opponent channels of CIELab space ("a","b") and the "L" channel.

2) Global Contrast: calculates the Global Contrast Factor [1] which measures contrasts at various resolution levels. Better corresponds to human contrast perception.

3) Colourfulness: metric that quantifies the perceived colourfulness of natural images [2]

4) Focus/Blur: measures the degree of focus of an image using a diagonal Laplacian (LAPD). Obtained from Matlab repository [3]. See online article [4]

5) Saturation (obtained from the HSV colour model)

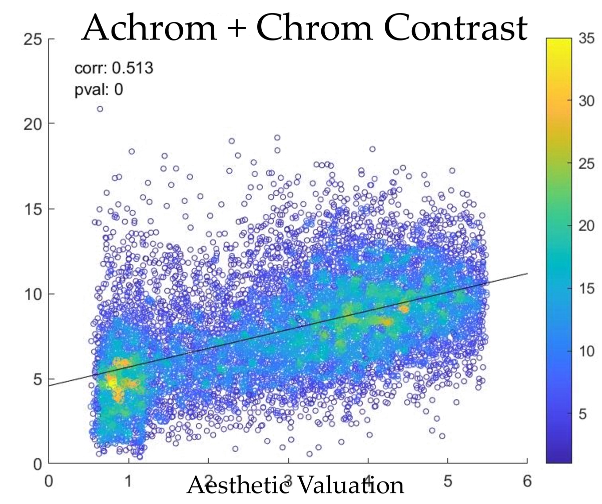
6) Mean grey-level value.

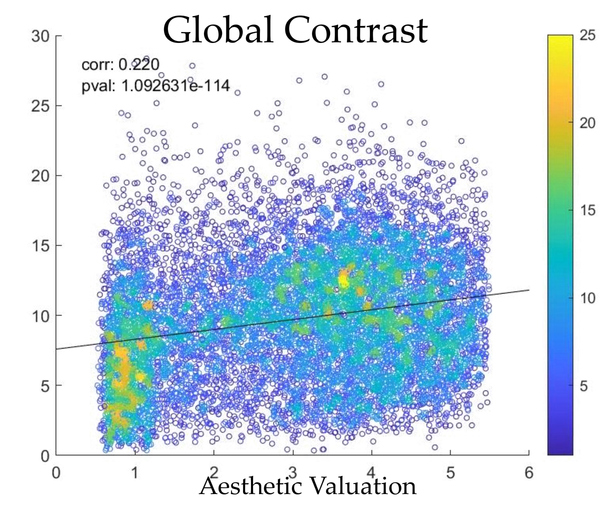
7) Mean lightness values (obtained from the CIELab "L" value.

8) Fourier "alpha" slope [5]

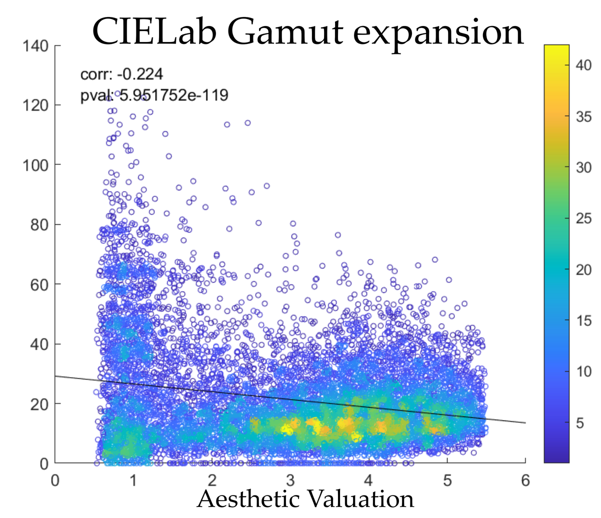
9) CIELab Gamut expansion: average distance of pixels to the white locus in CIELab colour space.

Non-learning measures	Semantic-Free		AVA	
Feature	Pearson's R	p-value	Pearson's R	p-value
Achrom+Chrom_Contrast	0.513	p <<	0.004	0.02
Global_Contrast	0.220	p <<	0.004	0.03
Colourfulness	-0.029	0.003	-0.008	p <<
Focus/Blur	-0.037	0.0001	0.064	p <<
HSV_Saturation	-0.088	p <<	0.002	0.3
Mean_Greylevel	-0.106	p <<	-0.032	p <<
Mean_Lightness (CIELab)	-0.135	p <<	-0.034	p <<
Fourier_Alpha_Slope	-0.209	p <<	-0.174	p <<
CIELab_Gamut_Expansion	-0.224	p <<	-0.017	p <<





Learning-measures	Semantic-Free		AVA	
Feature	Pearson's R	p-value	Pearson's R	p-value
Positive_Sentiment	0.397	p <<	0.026	p <<
Neutral_Sentiment	-0.106	p <<	0.094	p <<
Mean Depth	-0.188	p <<	-0.085	p <<
Negative_Sentiment	-0.275	p <<	-0.118	p <<



We compared the results obtained in our semantics-deprived dataset with one of the traditional datasets used in computational aesthetics (the AVA [6]) dataset using the features described above. The correlations obtained are shown in the table above. The figures on the right show the plots of the three largest correlations (Chromatic Contrast, Global Contrast and Colourfulness).

Additionally, we explored four extra features obtained using machine learning techniques:

10) Sentiments: this metric quantifies the sentiments elicited by an image, using a neural network [7] trained with 3 million tweets and images. The sentiments are classified into positive, negative and neutral.

11) Mean depth: tries to quantify the mean depth of the image using a neural network called "Vision Transformers for Dense Prediction". Images in the dataset were processed to have a maximum size of 150 (to save computer power)

The results show some correlation with the positive sentiment measure, which might be due to the role that contrast and gamut expansion plays on this metric. Mean depth also seems to be able to predict part of the aesthetic valuation results.

Conclusions

Our results show that the strongest candidate to explain aesthetic valuation of images is the "Chromatic contrast" (r= 0.513), followed by "Global Contrast" (r= 0.220), "Fourier Alpha Slope" (r= -0.209) and "Gamut Expansion" (r= -0.224). For comparison, we run the same analysis using a more traditional aesthetics image database (AVA, CVPR2012) with inconclusive or negative results. We also tested some features obtained the application of machine learning techniques. In summary our semantic-deprived dataset shows correlations that other datasets hide.

References:

[1] Krešimir Matkovic, László Neumann, Attila Neumann, Thomas Psik and Werner Purgathofer, Global Contrast Factor - a New Approach to Image Contrast, in Computational Aesthetics in Graphics, Visualization and Imaging (2005), L. Neumann, M. Sbert, B. Gooch, W. Purgathofer (Eds)

[2] Hasler, D., & Suesstrunk, S. E. (2003). Measuring colorfulness in natural images. Human Vision and Electronic Imaging VIII. doi:10.1117/12.477378

[3] www.mathworks.com/matlabcentral/fileexchange/27314-focus-measure

[4] https://arxiv.org/ftp/arxiv/papers/1604/1604.00546.pdf

[5] D. J. Tolhurst, Y. Tadmor and Tang Chao, Amplitude spectra of natural images, Ophthal. Physiol. Opt., 1992, Vol. 12.

[6] Naila Murray, Luca Marchesotti and Florent Perronnin, AVA: A Large-Scale Database for Aesthetic Visual Analysis, CVPR2012

[7] Lucia Vadicamo, Fabio Carrara, Andrea Cimino, Stefano Cresci, Felice Dell'Orletta, Fabrizio Falchi, Maurizio Tesconi, Cross-Media Learning for Image Sentiment Analysis in the Wild, ICCV2017

[8] R. Ranftl, A. Bochkovskiy and V. Koltun, "Vision Transformers for Dense Prediction," 2021 IEEE/CVF ICCV 2021

Acknowledgements

CAP and XO are partially supported by the Spanish Ministry of Science and Innovation through research project (Ref. PID2020-118254RB-I00)