Learning beautiful (and ugly) attributes

Luca Marchesotti
luca.marchesotti@xerox.com
Florent Perronnin
florent.perronnin@xerox.com

XRCE
Xerox Research Centre Europe
Meylan, France

The problem. The amount of visual content we handle on a daily basis has grown exponentially. In this ocean of images and videos, there are many questions that artificial systems could help us answer. In the last decade, the focus of the computer vision community had been on semantic recognition. While this is still a very active research field, new questions are arising. For instance, we might want to predict what people like in an image or a video. Although this is a very challenging question, even for humans, it was shown experimentally that aesthetics/preference can be predicted using data-driven approaches [1, 2]. Current approaches to aesthetic image analysis either provide accurate or interpretable results. We thus raise the following question: *can we preserve the advantages of generic features* ***and*** *get interpretable results?*.

The solution. In this work, we will address this problem by *discovering and learning attributes automatically*. In particular, our main contribution is a novel approach to aesthetic image analysis which combines the benefits of "attribute-based" and "generic" techniques. It consists of (i) automatically discovering a vocabulary of visual attributes and (ii) learning their visual appearance using generic features. For this purpose, we leverage the AVA dataset [4] which contains more than 250,000 images together with their aesthetic preference ratings and textual comments. Preference ratings allow us to supervise the creation of the attribute vocabulary (step (i)) and to learn automatically the visual appearance of attributes (step (ii)). Finally we apply these learned attributes to three different scenarios: aesthetic quality prediction, image classification and retrieval.

The data. We use AVA, a recently introduced database [4] which contains more than 250,000 images downloaded from www.dpchallenge.com. An interesting characteristic of this dataset is that images are accompanied by natural language text and attractiveness scores. This dataset was assembled for large-scale evaluation of attractiveness classification and regression tasks. But it was also recently used to study the dependence of attractiveness on semantic information [3]. Another peculiarity of this corpus is the organization of photos in *contests*: an equivalent of Flickr groups where images are ranked according to attractiveness scores left by users. Consider the sample images in figure 1, they were taken from the contest *"**Green Macro**: Get up close and personal with the subject of your choice, using green as your primary color"*. Photos in the first row scored highly, the others were ranked at the bottom of the contest. While all six images contain a lot of green, the top ones have brighter, more vivid green elements and the photographic technique *"Macro"* is much better represented.

Discovering aesthetic attributes. Mining attributes by hand-picking photographic rules from a book is problematic: this is a non-exhaustive procedure and it does not give any indication of how much, and when, these techniques should be used. Therefore, we mine beautiful and ugly attributes by discovering which terms can predict the aesthetic score of an image from AVA textual comments. For this purpose, we train an Elastic Net to predict aesthetic scores and, at the same time, select textual features. It is a regularized regression method that combines an ℓ_2-norm and a sparsity-inducing ℓ_1-norm. Let N be the number of textual documents. Let D be the dimensionality of the BOW histograms. Let $\mathbf{X}$ be the $N \times D$ matrix of documents. Let y be the $N \times 1$ vector of scores of aesthetic preference (the score of an image is the average of the scores it received). We learn:

$$\hat{\beta} = \arg\min_{\beta} ||\mathbf{y} - \mathbf{X}\beta||^2 + \lambda_1 ||\beta||_1 + \lambda_2 ||\beta||^2 \quad (1)$$

where λ_1 and λ_2 are the regularization parameters.

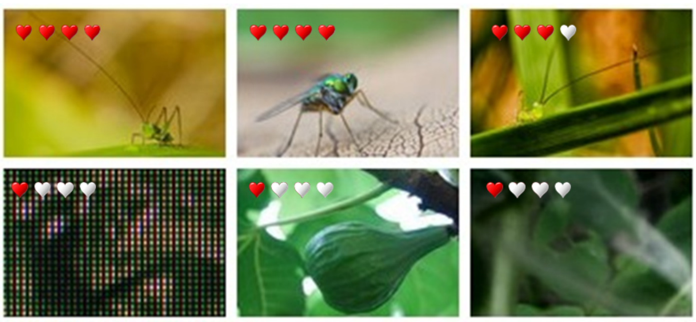

Figure 1: Sample photos from the challenge "Green Macro"

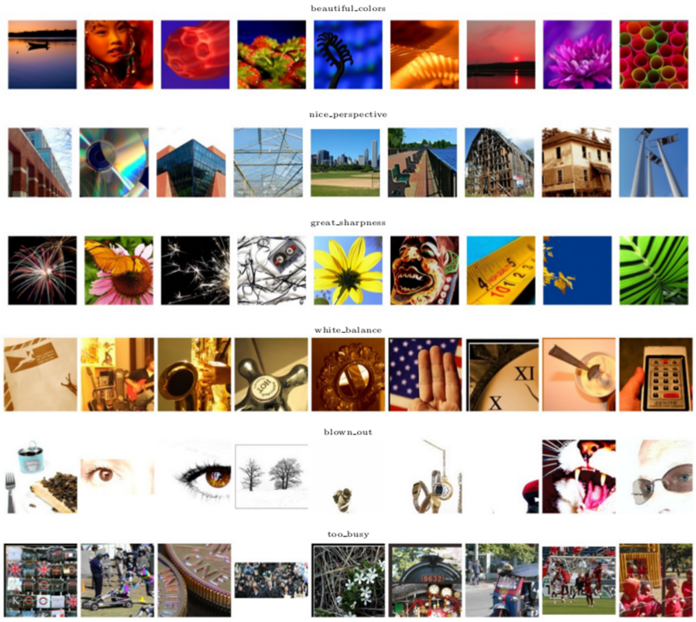

Figure 2: Images with top scores for some representative beautiful and ugly attributes. From top to bottom: beautiful colors, nice perspective, great sharpness, color cast, white balance, blown out, too busy.

Assessing attributes visualness. We enforced interpretability and discriminability of the attribute labels using attractiveness scores as a supervision mechanism. However, this choice does not ensure that all these attributes can be recognized by a computer. This is the reason why we measure "visualness" using Area Under the ROC Curve (AUC) calculated for each individual attribute. In particular, we benchmark the classification performances of each attribute (1-vs-all) and we rank them using AUC. We show some qualitative results in Figure 2 for Ugly and Beautiful attributes.

Conclusions. In this paper, we tackled the problem of visual attractiveness analysis using visual attributes as mid-level features. Despite the great deal of subjectivity of the problem, we showed that we can learn automatically meaningful attributes that can be used in various applications such as score prediction, auto-tagging or retrieval. Future work will focus on testing with users the advantage of our beautiful and ugly attributes and on mitigating biases introduced by semantic information.

[1] R. Datta, D. Joshi, J. Li, and J.Z. Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV*, 2006.

[2] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *CVPR*, 2006.

[3] N. Murray, L. Marchesotti, and F. Perronnin. Learning to rank images using semantic and aesthetic labels. In *BMVC*, 2012.

[4] N. Murray, L. Marchesotti, and F. Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *CVPR*, 2012.