Emotional Attention: A Study of Image Sentiment and Visual Attention (Supplementary Material)

Shaojing Fan[1], Zhiqi Shen[1], Ming Jiang[2], Bryan L. Koenig[3], Juan Xu[2], Mohan S. Kankanhalli[1], and Qi Zhao[2]

[1]National University of Singapore [2]University of Minnesota [3]Southern Utah University

This supplementary material provides additional details and examples to complement the main paper.

1. Construction of EMOd dataset

In this section, we provide more details on the EMOd dataset construction, including additional information on image collection, eye-tracking experiment, and image annotation.

1.1. Image collection

The EMOd dataset was constructed from two sources: (1) a subset (321) photos of the International Affective Picture System (IAPS) [9]; (2) a set of 698 photos collected by the authors. For IAPS, we used a selection of 321 photos each identified as primarily eliciting one emotion in a study by [12]. This subset has also been used in other computer vision research on emotion assessment [16, 11, 10]. The aim of our own collection was to make the dataset more diverse regarding how observers' emotions are evoked. We grouped the 698 images into six types based on how they evoked emotions (parenthetical numbers are how many images were that type): emotion-eliciting objects (29), emotion-eliciting activities (158), emotion-eliciting gist (145), emotion-eliciting spatial layout (105), emotion-eliciting color and illumination (121), and emotionally-neutral images (140). Fig. 1 shows example images of the six types.

1.2. Psychophysics study I: eye tracking

Here we report detailed settings of the eye-tracking experiment.

Sixteen subjects freely observed all images in the EMOd dataset. The subjects were undergraduate and graduate students, ages ranging from 21 to 35 years old (27.0 ± 4.7). All subjects had normal or corrected-to-normal visual acuity. Images were presented on a 22-inch LCD monitor. The screen resolution was 1920×1080. Images were scaled to occupy the screen's full height with the aspect ratio of the images unchanged. The visual angle of the stimuli was about $38.94° \times 29.20°$, and each degree of visual angle contained about 26.3 image pixels. Subject eye movements were recorded at 1000Hz using an Eyelink 1000 eye tracker. The study was split into three sessions, with 389, 430, and 430 randomly ordered images respectively, and each session was completed within one hour. A 9-point calibration was performed before each recording session. Each image was presented for 3 seconds, followed by a drift correction that required subjects to fixate in the center of the screen and press the space bar to continue.

1.3. Psychophysics study II: image annotation

We built an online EMOd object-labeling system based on the LabelMe platform [13], and an online EMOd image-annotation platform. Fig. 2 shows the user interfaces of the two platforms.

Each object was labeled according to its sentiment category (either negative, neutral, or positive) and semantic category. The design of semantic categories is based on [15], which includes four types: (1) directly relating to humans (*i.e.*, emotional face, neutral face, touched, gazed), (2) relating to other (nonvisual) senses of humans (*i.e.*, sound, smell, taste, touch), (3) designed to attract attention or for interaction with humans (*i.e.*, text, watchability, operability), (4) objects with implied motion. Fig. 3 shows example images containing objects of each semantic category.

Each image was further rated on 33 high-level perceptual attributes on a Likert scale of 1 (not at all) to 9 (very much). We designed an attributes list covering both semantic and sentiment aspects of the image, including (1) 10 basic emotions commonly studied in psychology [4, 12]: happiness, surprise, awe, excitement, amusement, contentment, sadness, anger, fear, and disgust; (2) Self-Assessment Manikin for non-verbal pictorial assessment [1]: valence, arousal, dominance; (3) high-level attributes commonly studied in

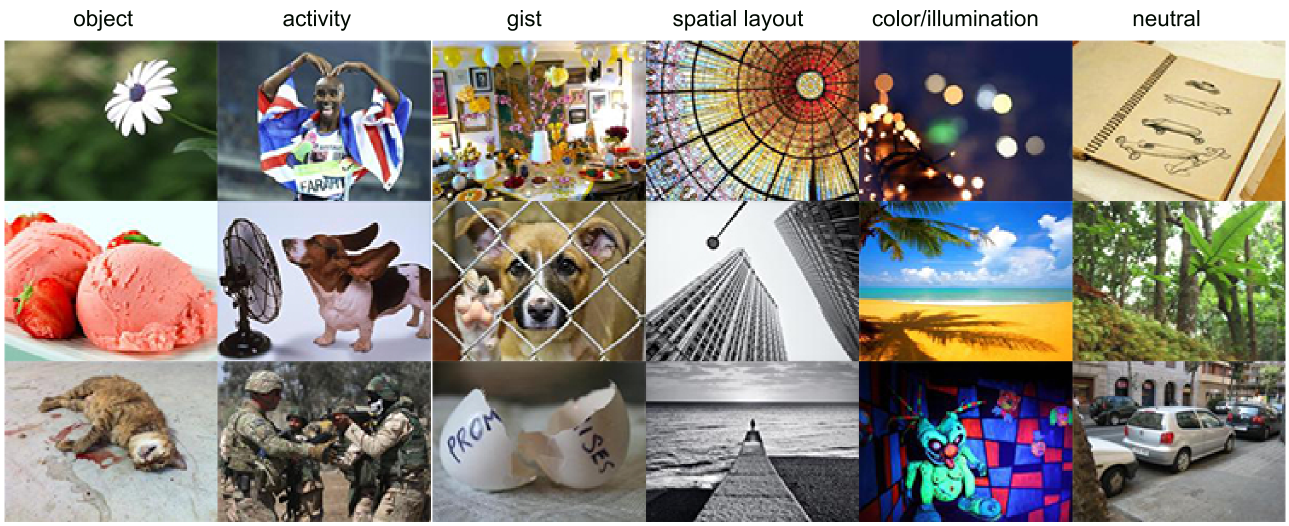

Figure 1: Example images from EMOd dataset illustrating the types based on how they evoked emotions in observers.

(a) (b)

Figure 2: User interface of (a) EMOd object-labeling platform, and (b) EMOd image-annotation platform.

computer vision, such as aesthetics, image quality, photo-realism, depths of field, and symmetry [7, 10, 6]. Table 1 shows the detailed list of the 33 attributes.

For the 698 images we collected, we deployed the E-MOd image-annotation platform on Amazon Mechanical Turk (AMT) [2] and recruited 348 AMT workers ($>95\%$ approval rate in Amazon's system) to annotate. For the IAPS data set, due to copyright restrictions, we recruited 10 undergraduate students to annotate them on the platform within the campus intranet. On average, each image was annotated by 10 participants. For each image we computed the score of each attribute by averaging the answers given by the 10 participants, then transformed scores for each attribute to a range of [0, 1] with raw scores of 1 becoming 0 and raw scores of 9 becoming 1. Each image was further classified into one of the following 8 scene categories [14] by two paid undergraduate students (numbers in parentheses are the number of images in each category): human (363), animal (117), architecture (105), vehicle (65), natural scenery (145), static object (123), urban (63), and indoor (59).

References

[1] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994. 1, 3

[2] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011. 2

[3] R. Datta, J. Li, and J. Z. Wang. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In *ICIP*, pages 105–108. IEEE, 2008. 3

[4] P. Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992. 1, 3

[5] S. Fan, T.-T. Ng, J. S. Herberg, B. L. Koenig, C. Y.-C. Tan, and R. Wang. An automated estimator of image visual realism based on human cognition. In *CVPR*, pages 4201–4208. IEEE, 2014. 3

[6] S. Fan, T.-T. Ng, B. L. Koenig, M. Jiang, and Q. Zhao. A paradigm for building generalized models of human image perception through data fusion. In *CVPR*, pages 5762–5771, 2016. 2, 3

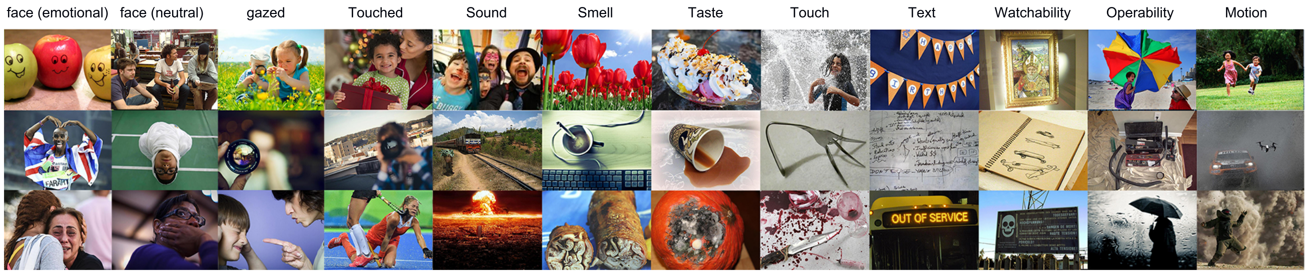

Figure 3: Example images from EMOd dataset illustrating semantic categories. Each column is a list of objects with each semantic category.

Attribute type	**Detailed attributes**
Emotions [4, 12]	Happiness; Surprise; Awe; Excitement; Amusement; Contentment; Sadness; Anger; Fear; Disgust
Self-Assessment Manikin [1]	Valence; Arousal; Dominance
Semantics [6]	Familiarity; Unusualness; Dynamics; Informativeness; Natural object combination
Aesthetics [8, 3]	Aesthetics; High quality; Colorfulness; Natural color combination; Sharpness
Spatial layout [10]	Have objects of focus; Single object focus; Close-up shot; Centered; Symmetry
Naturalness [5]	Photorealism
Related to people [5]	Attractive person; Posing; Eye contact; Positive expression

Table 1: List of 33 high-level perceptual attributes in the EMOd dataset.

[7] P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. In *NIPS*, pages 2429–2437, 2011. 2

[8] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *CVPR*, volume 1, pages 419–426. IEEE, 2006. 3

[9] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (iaps): Affective ratings of pictures and instruction manual. *Technical report A-8*, 2008. 1

[10] X. Lu, P. Suryanarayan, R. B. Adams Jr, J. Li, M. G. Newman, and J. Z. Wang. On shape and the computability of emotions. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 229–238. ACM, 2012. 1, 2, 3

[11] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM Multimedia*, pages 83–92. ACM, 2010. 1

[12] J. A. Mikels, B. L. Fredrickson, G. R. Larkin, C. M. Lindberg, S. J. Maglio, and P. A. Reuter-Lorenz. Emotional category data on images from the international affective picture system. *Behavior research methods*, 37(4):626–630, 2005. 1, 3

[13] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1-3):157–173, 2008. 1

[14] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR*, pages 3485–3492. IEEE, 2010. 2

[15] J. Xu, M. Jiang, S. Wang, M. S. Kankanhalli, and Q. Zhao. Predicting human gaze beyond pixels. *Journal of vision*, 14(1):28–28, 2014. 1

[16] V. Yanulevskaya, J. Van Gemert, K. Roth, A.-K. Herbold, N. Sebe, and J.-M. Geusebroek. Emotional valence categorization using holistic image features. In *2008 15th IEEE International Conference on Image Processing*, pages 101–104. IEEE, 2008. 1