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

Shaojing Fan¹, Zhiqi Shen¹, Ming Jiang², Bryan L. Koenig³, Juan Xu², Mohan S. Kankanhalli¹, and Qi Zhao²

¹National University of Singapore  ²University of Minnesota  ³Southern 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° × 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 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
For the 698 images we collected, we deployed the EMOd 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

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

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

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