A Mixed Bag of Emotions: Model, Predict, and Transfer Emotion Distributions Supplementary Material

Kuan-Chuan Peng, Tsuhan Chen
Cornell University
{kp388, tsuhan}@cornell.edu

Amir Sadovnik
Lafayette College
sadovnia@lafayette.edu

Andrew Gallagher
Google Inc.
agallagher@google.com

1. Images in Emotion6

The following subsections display the images in Emotion6 according to the dominant evoked emotion, the category among anger, disgust, fear, joy, sadness, surprise, and neutral with the highest probability selected as the evoked emotion from user study.

1.1. Anger

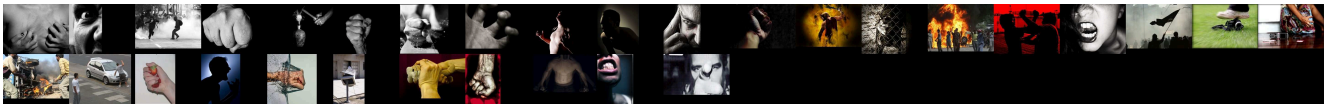

Figure 1. The images in Emotion6 with anger selected as the dominant evoked emotion.

1.2. Disgust

Figure 2. The images in Emotion6 with disgust selected as the dominant evoked emotion.

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

1.3. Fear

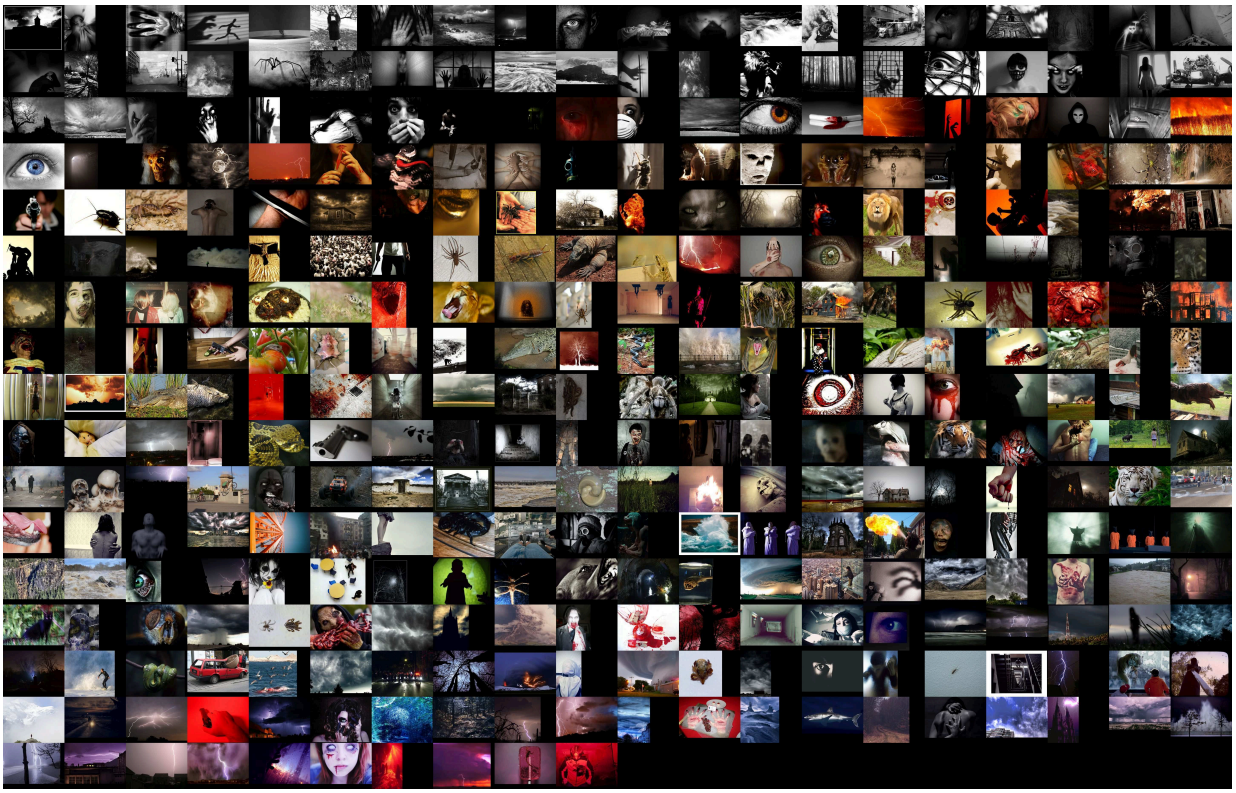

Figure 3. The images in Emotion6 with fear selected as the dominant evoked emotion.

1.4. Sadness

Figure 4. The images in Emotion6 with sadness selected as the dominant evoked emotion.

1.5. Joy

Figure 5. The images in Emotion6 with joy selected as the dominant evoked emotion.

1.6. Surprise

Figure 6. The images in Emotion6 with surprise selected as the dominant evoked emotion.

1.7. Neutral

Figure 7. The images in Emotion6 with neutral selected as the dominant evoked emotion.

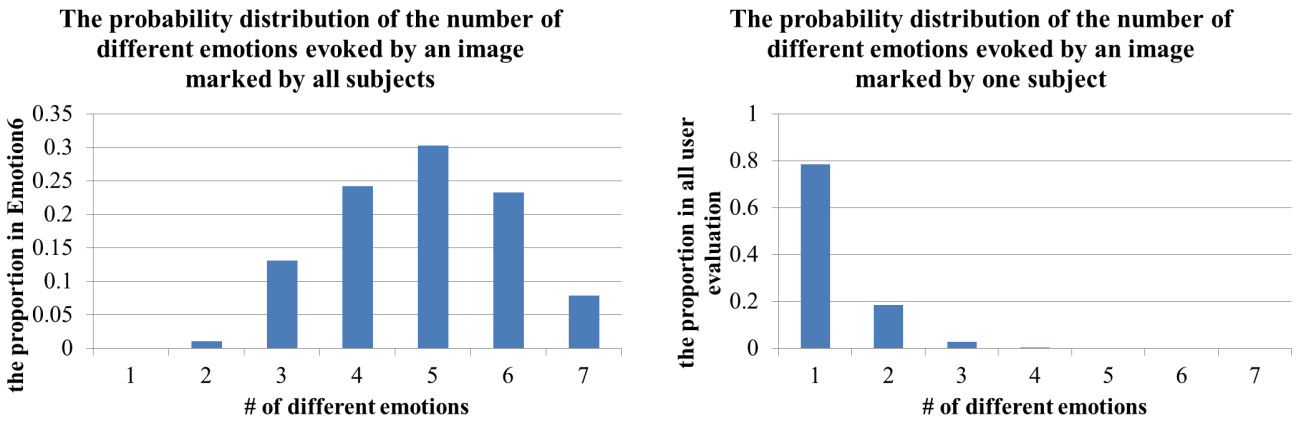

Figure 8. The probability distribution of the number of emotions evoked by an image marked by all subjects viewing the image (left) and each subject (right).

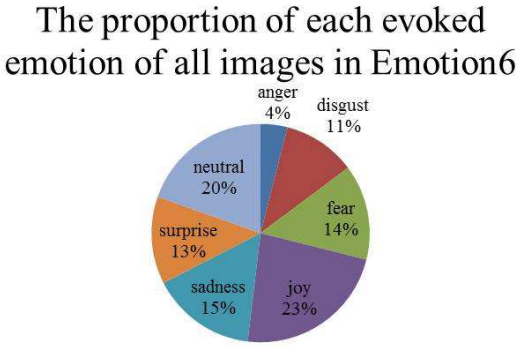

Figure 9. The proportion of each evoked emotion of all images in Emotion6.

2. Emotion6 Statistics

2.1. Emotion Keywords

Collecting images for Emotion6 by typing Ekman's 6 basic emotions [2] (anger, disgust, joy, fear, sadness, and surprise) and their synonyms as searching keywords on Flickr, we put these images (330 images for each searching keyword) on Amazon Mechanical Turk (AMT) and ask the subjects to select keyword(s) as the evoked emotion of each image. The subjects can choose one or more keywords from 7 candidates we provide (Ekman's 6 basic emotions [2] and neutral) to describe the evoked emotion of the image.

First, we calculate the probability distribution of the number of emotion categories evoked by a single image in Emotion6. The result in Figure 8 verifies the existence of the mixed bag of emotions, and most images in Emotion6 can evoke emotions from at least 4 categories. Second, the proportion of each emotion keyword marked as evoked emotion is shown in Figure 9. Even though we collect the same number of images for each category, Figure 9 shows uneven distribution in terms of emotion keywords because the keyword used to search an image may not be the most representative one. This is ignored in constructing previous databases like emodb [7] and the "artistic photographs" database [4]. Among all 7 categories, anger is the one selected the least, indicating that images tagged with anger in Emotion6 often do not convey or elicit anger. Third, we calculate the proportion of each emotion keyword marked as the evoked emotion given a certain keyword used to search the images. The results in Figure 10 show the noticeable difference between the searching keywords and evoked emotions. Figure 10 also shows the proportion of each emotion keyword used to search the images given a certain evoked emotion.

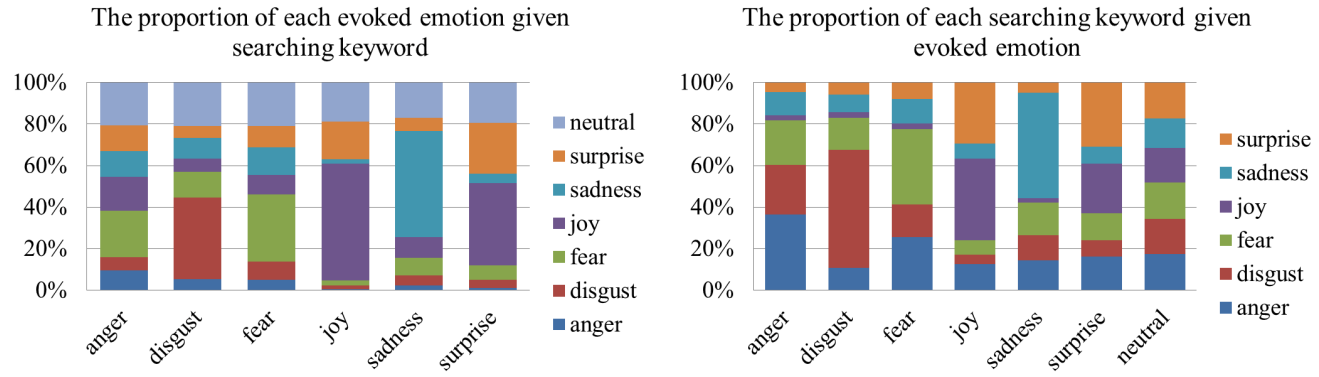

Figure 10. The left graph shows the proportion of evoked emotion given each searching keyword in Emotion6. The right graph shows the proportion of searching keywords given each evoked emotion in Emotion6.

Figure 11. The mapping between the color codes and emotion keywords used in Section 2.2.

2.2. Valence–Arousal (VA) Scores

In this section, we present statistics related to VA scores in SAM 9-point scale [1]. For the valence scores, 1, 5, and 9 mean very negative, neutral, and very positive emotions respectively. For the arousal scores, 1 (9) means the emotion has low (high) stimulating effect. The boundary of each image is colored according to its dominant evoked emotion using the color codes in Figure 11. Figure 12 places all the images of Emotion6 in VA plane according to the ground truth of evoked VA scores. We also show some images in Emotion6 with extreme standard deviation of V and A (σ_V and σ_A), which means the subjects have different degree of consensus for each image in terms of VA scores. Figure 13 and Figure 14 display images with extreme σ_V and σ_A respectively with the ground truth VA scores, σ_V and σ_A in evoked emotion.

We also show the relationship between emotion keywords and VA scores in Figure 15, where the color of each point represents the probability of all the subjects giving certain VA scores given they choose the specified keyword as the evoked emotion. The distribution in general meets our expectation and reflects the correlation between VA scores and emotion keywords. For example, surprise has higher arousal scores and joy has higher valence scores. However, the distribution using images as input is different from that using keywords as input [3] because people have more consensus in interpreting words than images. Figure 15 also shows that the areas occupied by each emotion keyword in VA space are different but can overlap, which means VA scores cannot totally replace emotion keywords. Since VA scores and emotion keywords capture different aspects of emotions, we predict both VA scores and emotion distribution.

3. Predict Valence–Arousal (VA) Scores

Method 1	Method 2	P_V	P_A
CNNR	Popularity	0.631	0.577
CNNR	Random	0.729	0.818
CNNR	SVR	0.556	0.502

Method	AAD of Valence	AAD of Arousal
Popularity	1.590	0.829
Random	2.423	2.113
SVR	1.347	**0.734**
CNNR	**1.219**	0.741

Table 1. The performance of different algorithms for predicting valence and arousal scores. The numbers are the average of absolute difference (AAD) compared with the ground truth in SAM 9-point scale [1]. CNNR outperforms the two baseline, and has comparable performance to SVR.

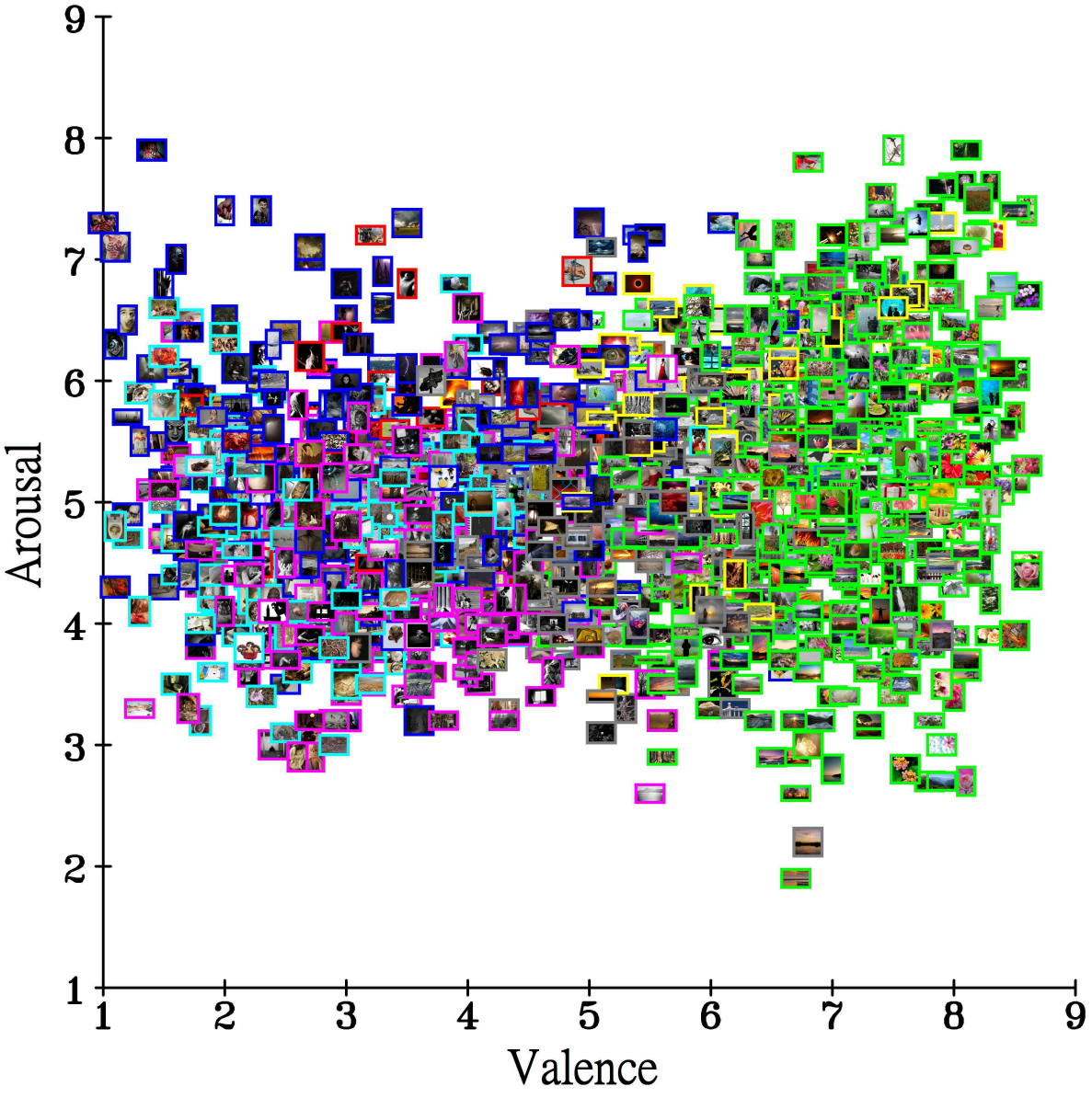

Figure 12. The distribution of evoked VA scores of Emotion6. All the images in Emotion6 are placed in VA plane according to their evoked VA scores. The boundary of each image is colored to reflect the dominant evoked emotion according to the color codes in Figure 11.

We create predictors for VA scores separately using the same set of features and similar methods as those of predicting the emotion distribution in the main paper. We compare the results of SVR and CNNR with two baselines: 1: Guessing V (A) score as the mode of all V (A) scores. 2: Guessing VA scores uniformly. We evaluate the results with the average of absolute difference (AAD) compared with the ground truth in SAM 9-point scale [1]. We also report P_V (P_A), the proportion of the images in the test set where Method 1 predicts more accurate V (A) than Method 2. For the baseline using uniform random guessing, we repeat 100000 times and report the average. The results are listed in Table 1. CNNR outperforms the two baselines, and has comparable performance with respect to SVR.

Figure 13. The images in Emotion6 with extreme σ_V (marked in bold) in evoked emotion. The top (bottom) row shows the images in Emotion6 with the lowest (highest) σ_V in evoked emotion. The ground truth of evoked VA scores, σ_V, and σ_A is provided under each image. The boundary of each image is colored to reflect the dominant evoked emotion according to the color codes in Figure 11.

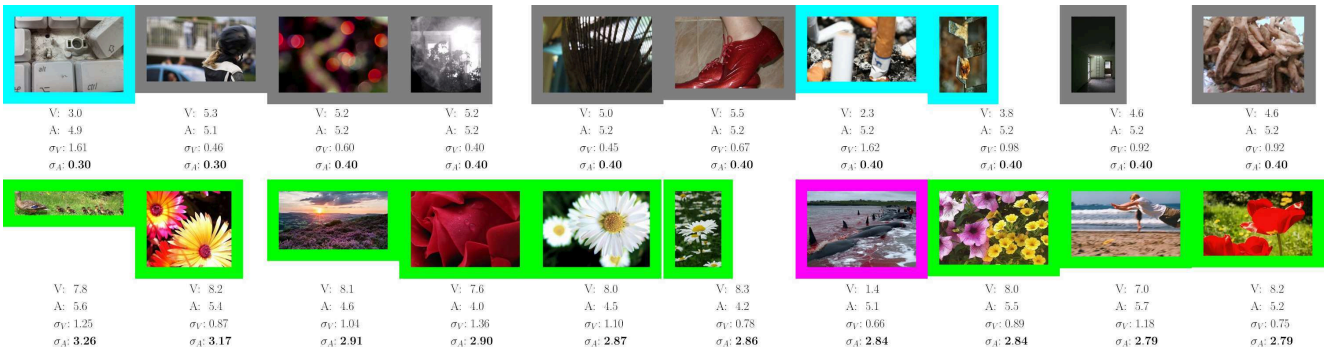

Figure 14. The images in Emotion6 with extreme σ_A (marked in bold) in evoked emotion. The top (bottom) row shows the images in Emotion6 with the lowest (highest) σ_A in evoked emotion. The ground truth of evoked VA scores, σ_V, and σ_A is provided under each image. The boundary of each image is colored to reflect the dominant evoked emotion according to the color codes in Figure 11.

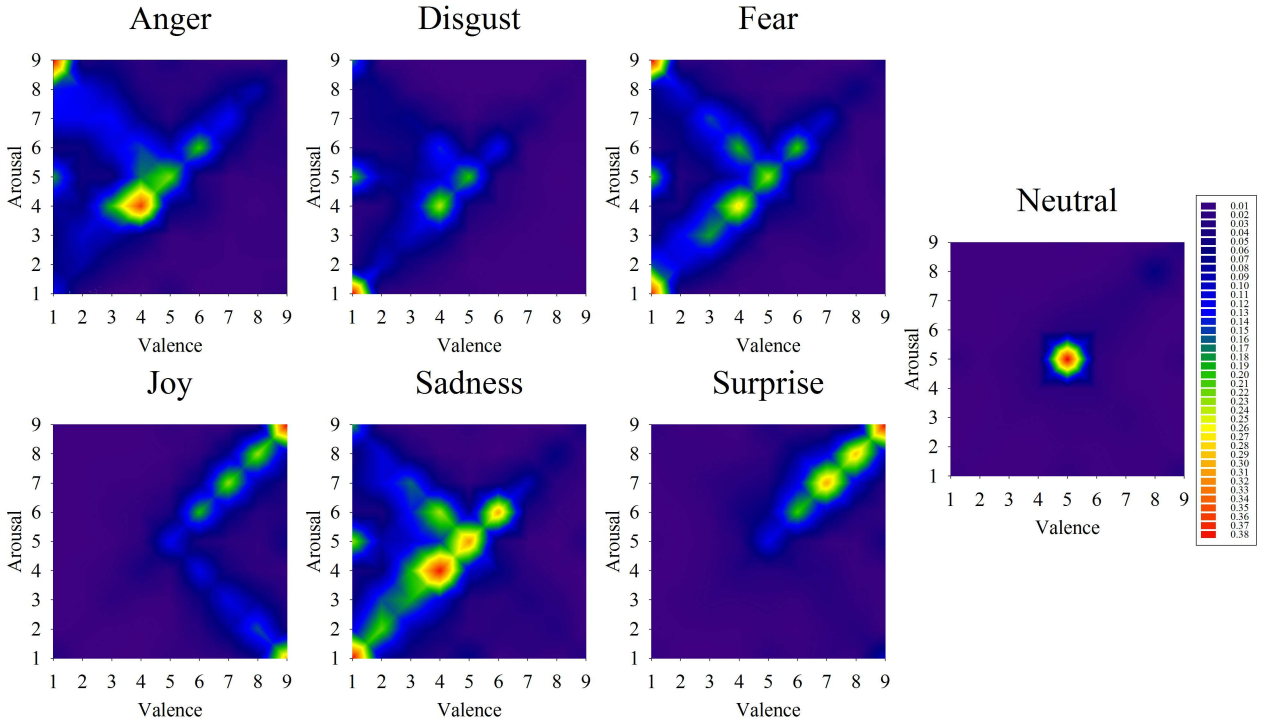

Figure 15. The probability of certain VA scores given each evoked emotion. The distribution reflects the correlation between VA scores and emotion keywords with images as input.

4. Examples of Emotion Transfer

The examples in Figure 16, 17, and 18 show that we can adjust the evoked emotion distribution of the source image toward that of the target by modifying the color tone and texture related features of the source image. The ground truth evoked emotion distribution is shown under each image. We use four different distance metrics to evaluate the similarity between two evoked emotion distributions – KL-Divergence (KLD), Bhattacharyya coefficient (BC), Chebyshev distance (CD), and earth mover's distance (EMD) [5, 6]. Since KLD is not well defined when a bin has value 0, we use a small value $\varepsilon = 10^{-10}$ to approximate the values in such bins. In computing EMD in our paper, we assume that each of the 7 dimensions (Ekman's 6 basic emotions [2] and neutral) is such that the distance between any two dimensions is the same. For KLD, CD and EMD, lower is better. For BC, higher is better.

For each distance metric M ($M \in \{KLD, BC, CD, EMD\}$), we compute the distances between: 1: source and target images $D_{M_s} = M(d_s, d_t)$. 2: transformed and target images $D_{M_{tr}} = M(d_{tr}, d_t)$, where d_s, d_t, and d_{tr} are the ground truth probability distributions of evoked emotion of the source, target, and transformed images respectively. The results are reported in terms of D_{M_s}, $D_{M_{tr}}$, and P_M, where P_M represents the probability that d_{tr} is closer to d_t than d_s is, using metric M. We also show some typical failure modes of emotion transfer in Fig. 19. There are two main reasons for such failure cases: 1: d_s may be mostly caused by the high level semantics such that the modification in low-level features can hardly shape d_s closer to d_t. 2: d_t may be also mostly caused by the high level semantics such that copying the low-level features of the target cannot totally replicate its emotional stimuli.

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. 6, 7

[2] P. Ekman, W. V. Friesen, and P. Ellsworth. What emotion categories or dimensions can observers judge from facial behavior? *Emotion in the Human Face*, pages 39–55, 1982. 5, 9

[3] J. R. J. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth. The world of emotions is not two-dimensional. *Psychological Science*, 18(2):1050–1057, 2007. 6

[4] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the International Conference on Multimedia*, pages 83–92, 2010. 5

[5] K. Matsumoto, K. Kita, and F. Ren. Emotion estimation of wakamono kotoba based on distance of word emotional vector. In *NLPKE*, pages 214–220, 2011. 9

[6] E. M. Schmidt and Y. E. Kim. Modeling musical emotion dynamics with conditional random fields. In *ISMIR*, pages 777–782, 2011. 9

[7] M. Solli and R. Lenz. Emotion related structures in large image databases. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, pages 398–405, 2010. 5

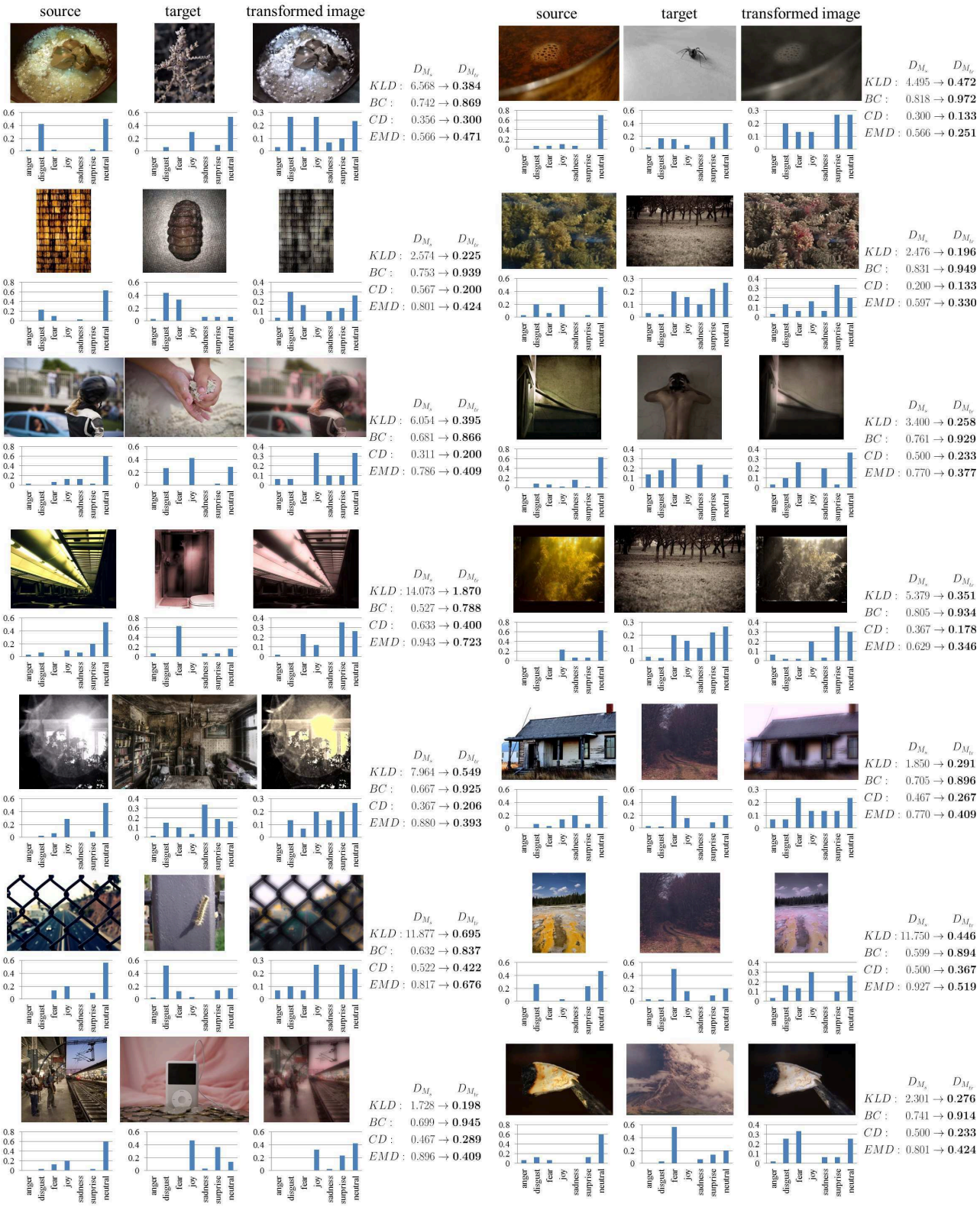

Figure 16. Examples showing the feature transform in transferring evoked emotion. For each example, D_{M_s} and $D_{M_{tr}}$ are provided ($M \in \{KLD, BC, CD, EMD\}$) with better scores marked in bold. The ground truth of evoked emotion distribution from AMT is provided under each image. In each example, the transformed image has closer evoked emotion distribution to that of the target compared to the source in all 4 metrics.

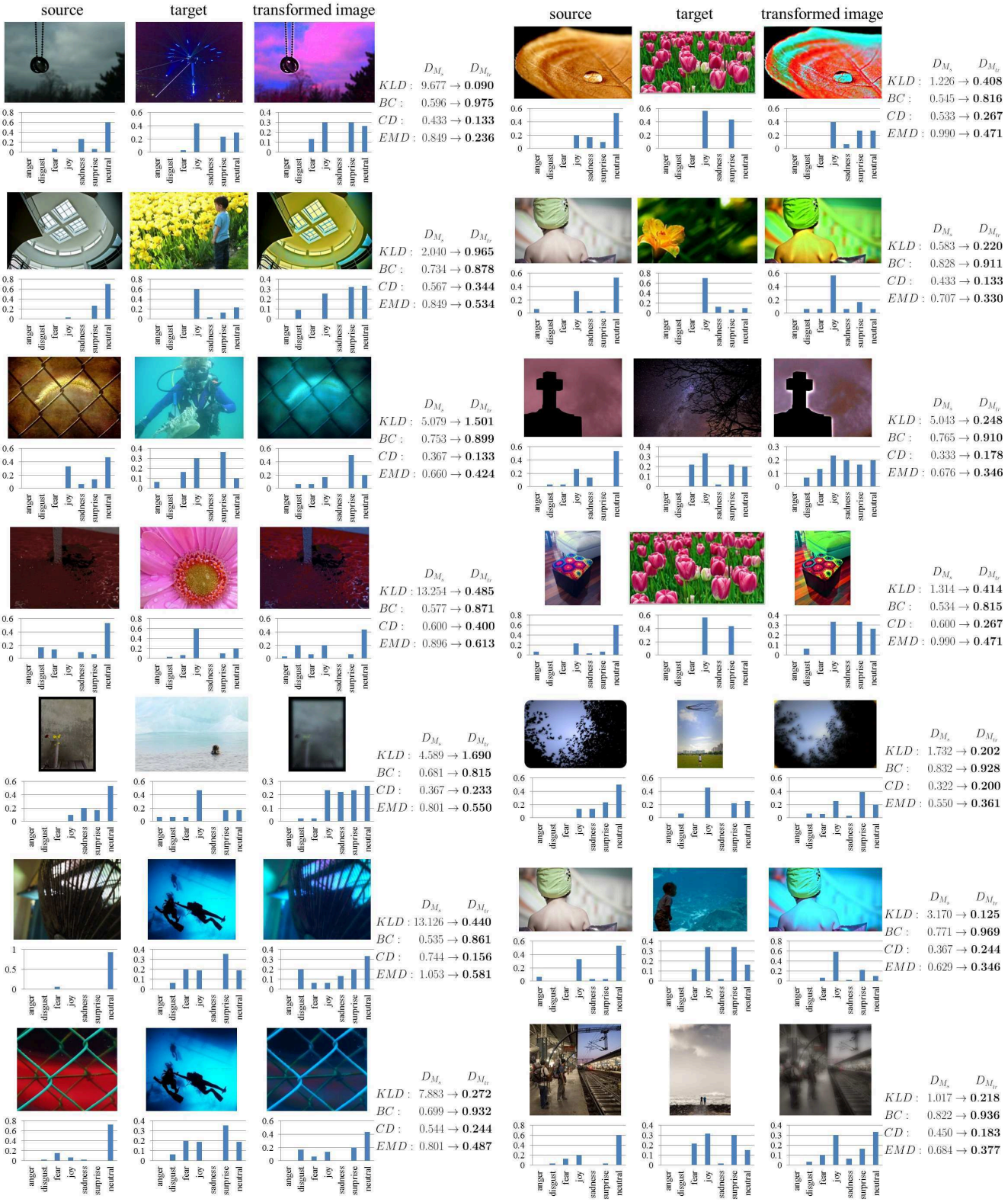

Figure 17. Examples showing the feature transform in transferring evoked emotion. For each example, D_{M_s} and $D_{M_{tr}}$ are provided ($M \in \{KLD, BC, CD, EMD\}$) with better scores marked in bold. The ground truth of evoked emotion distribution from AMT is provided under each image. In each example, the transformed image has closer evoked emotion distribution to that of the target compared to the source in all 4 metrics.

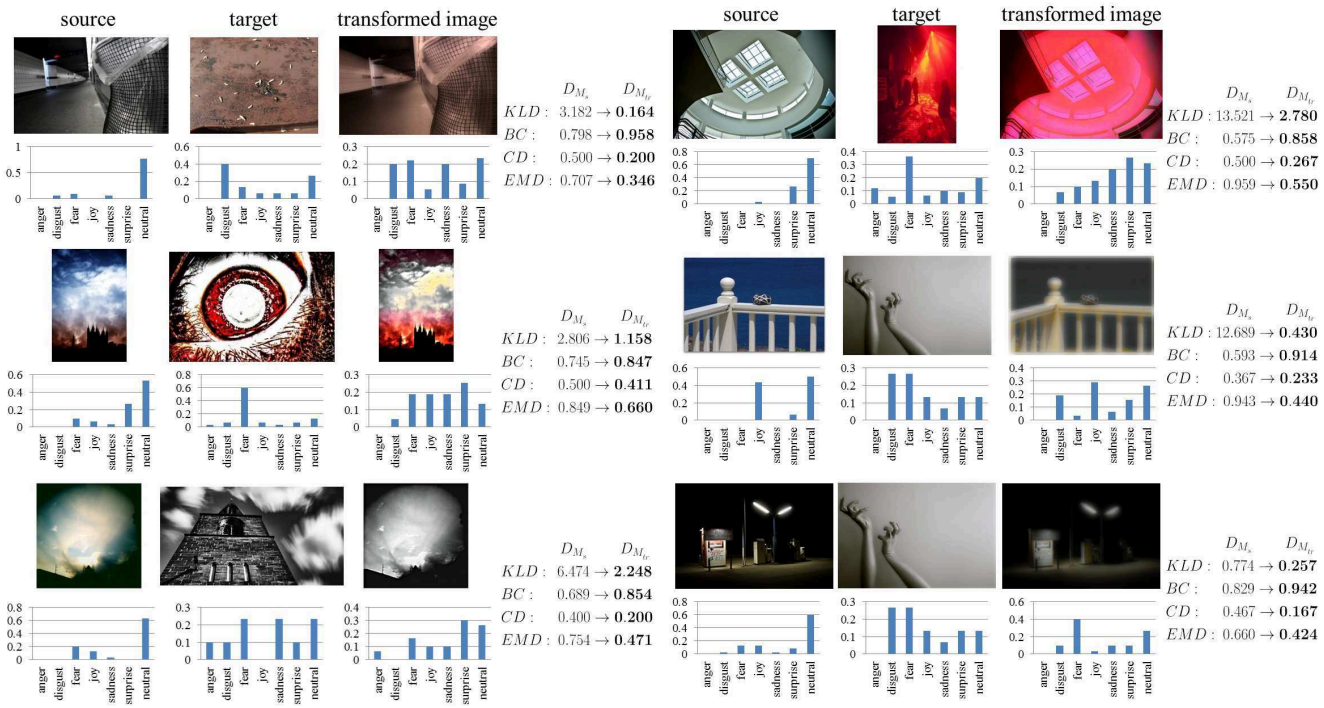

Figure 18. Examples showing the feature transform in transferring evoked emotion. For each example, D_{M_s} and $D_{M_{tr}}$ are provided ($M \in \{KLD, BC, CD, EMD\}$) with better scores marked in bold. The ground truth of evoked emotion distribution from AMT is provided under each image. In each example, the transformed image has closer evoked emotion distribution to that of the target compared to the source in all 4 metrics.

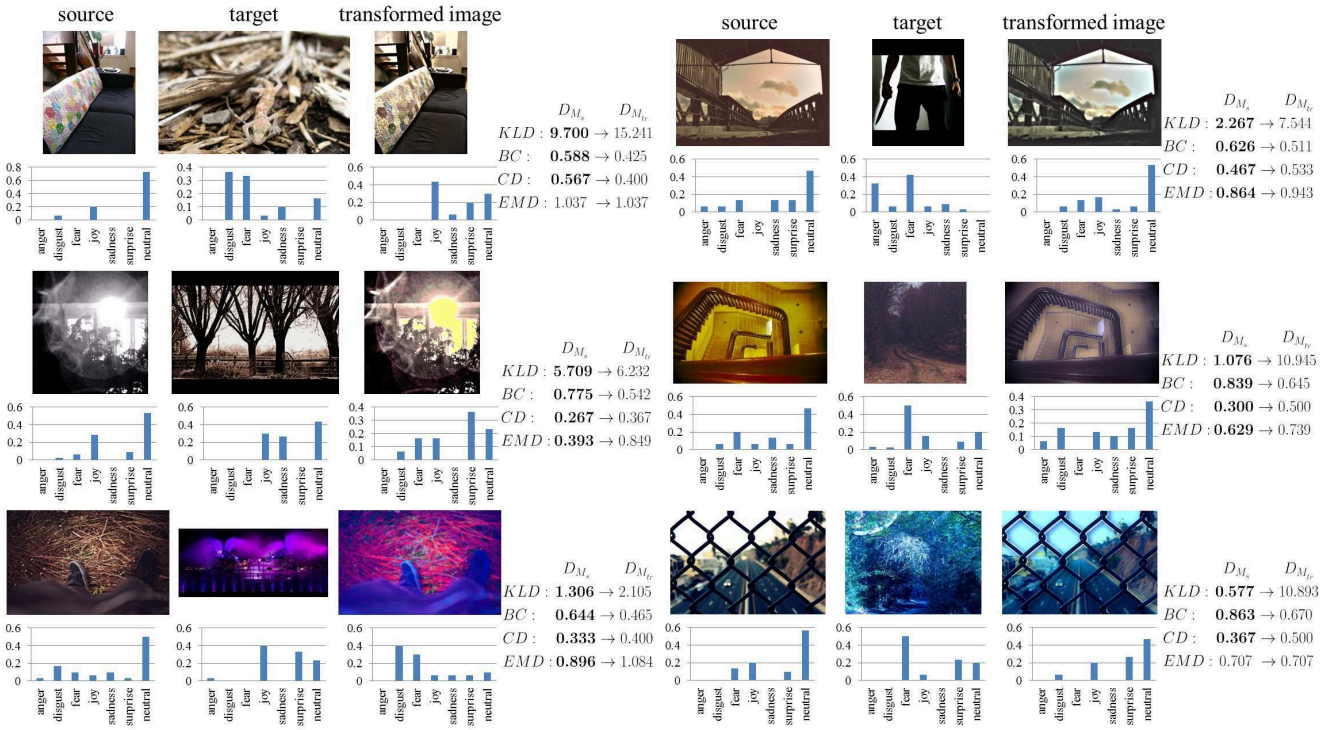

Figure 19. Failure examples of transferring emotions. The ground truth of evoked emotion distribution from AMT is provided under each image. For each example, D_{M_s} and $D_{M_{tr}}$ are provided ($M \in \{KLD, BC, CD, EMD\}$) with better scores marked in bold. The results show that the evoked emotion distribution of the source does not move toward that of the target in these examples.