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

Kuan-Chuan Peng[1], Amir Sadovnik[2], Andrew Gallagher[3], Tsuhan Chen[1]
[1]School of Electrical and Computer Engineering, Cornell University. [2]Department of Computer Science, Lafayette College. [3]Google Inc.

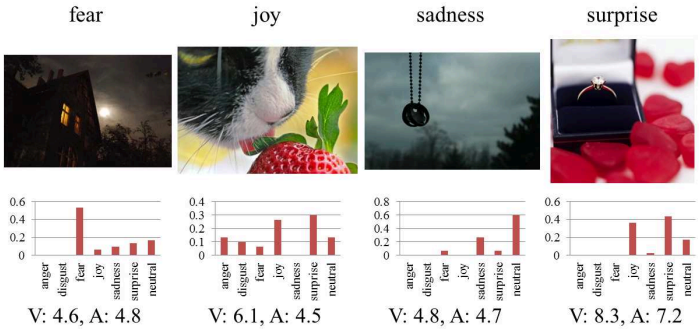

Figure 1: Example images of Emotion6 with the corresponding ground truth. The emotion keyword used to search each image is displayed on the top. The graph below each image shows the probability distribution of evoked emotions of that image. The bottom two numbers are valence–arousal (VA) scores in SAM 9-point scale [1].

This extended abstract summarizes our 3 contributions: 1) We show that different people have different emotional reactions to an image and that the same person may have multiple emotional reactions to an image. Our proposed database, Emotion6, addresses both findings by modeling emotion distributions. 2) We use a convolutional neural network (CNN) to predict emotion distributions, rather than simply predicting a single dominant emotion, evoked by an image. Our predictor of emotion distributions for Emotion6 is a better baseline than using support vector regression (SVR) with the features from previous works [3, 4, 5]. 3) We introduce the application of transferring the evoked emotion distribution from one image to another. With the support of a user study, we successfully adjust the evoked emotion distribution of an image toward that of a target image without changing the high-level semantics.

The Emotion6 Database: For each image in Emotion6, the following information is collected by a user study: 1) The ground truth valence–Varousal (VA) scores for evoked emotion. 2) The ground truth evoked emotion distribution. Figure 1 shows example images from Emotion6 with the corresponding ground truth.

Predicting Emotion Distributions: We compare our proposed 3 methods (SVR, CNN, and CNNR) with the 3 baselines (uniform, random, and optimally dominant (OD) distributions). Table 1 summarizes our proposed methods and the baselines. We use 4 different distance metrics to evaluate the similarity between two emotion distributions – KL-Divergence (KLD), Bhattacharyya coefficient (BC), Chebyshev distance (CD), and earth mover's distance (EMD). For KLD, CD and EMD, lower is better. For BC, higher is better. Table 2 summarizes the result of predicting emotion distributions.

Transferring Evoked Emotion Distributions: Our method adjusts the color tone and texture related features to modify the evoked emotion distribution of the source towards that of the target image. Fig. 2 is an example of emotion transfer using our method.

[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.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105. 2012.

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

[4] M. Solli and R. Lenz. Emotion related structures in large image databases. In *ACMCIVR*, pages 398–405, 2010.

[5] X. Wang, J. Jia, J. Yin, and L. Cai. Interpretable aesthetic features for affective image classification. In *ICIP*, pages 3230–3234, 2013.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

Method	Description
Uniform	A uniform distribution across all emotion categories.
Random	A random probability distribution.
OD	Optimally dominant (OD) distribution, a winner-take-all strategy where the emotion category with highest probability in ground truth is set to 1, and other emotion categories have zero probability.
SVR	We adopt features from [3, 4, 5] and train regressors using Support Vector Regression (SVR).
CNN	We adopt the CNN in [2] to predict the probability of the input image being classified as each emotion category.
CNNR	Similar to CNN except that the softmax loss layer is replaced with the Euclidean loss layer.

Table 1: The baseline methods (uniform, random, and OD) and our proposed methods (SVR, CNN, and CNNR) of predicting emotion distributions.

Method 1	Method 2	P_{KLD}	P_{BC}	P_{CD}	P_{EMD}
CNNR	Uniform	0.742	0.783	0.692	0.756
CNNR	Random	0.815	0.819	0.747	0.802
CNNR	OD	0.997	0.840	0.857	0.759
CNNR	SVR	0.625	0.660	0.571	0.620
CNNR	CNN	0.934	0.810	0.842	0.805
Uniform	OD	0.997	0.667	0.736	0.593

Method	$\overline{KLD}$	$\overline{BC}$	$\overline{CD}$	$\overline{EMD}$
Uniform	0.697	0.762	0.348	0.667
Random	0.978	0.721	0.367	0.727
OD	10.500	0.692	0.510	0.722
SVR	0.577	0.820	0.294	0.560
CNN	2.338	0.692	0.497	0.773
CNNR	**0.480**	**0.847**	**0.265**	**0.503**

Table 2: The performance comparison of predicting emotion distributions under P_M and $\overline{M}$ ($M \in \{KLD, BC, CD, EMD\}$). $\overline{M}$ is the mean of M, and P_M is the proportion of images where Method 1 matches the ground truth distribution more accurately than Method 2 according to distance metric M. CNNR performs the best out of all the listed methods in terms of all P_Ms with better $\overline{M}$.

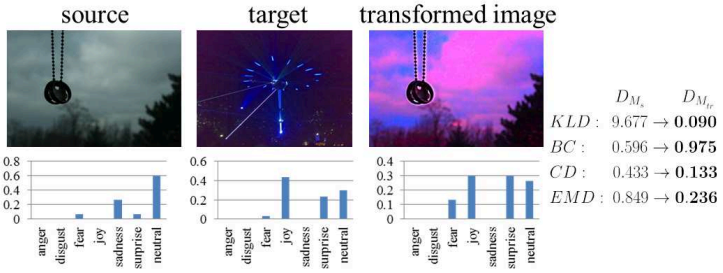

Figure 2: An example of transferring evoked emotion distribution. We transform the color tone and texture related features of the source to those of the target. The ground truth probability distribution of the evoked emotion is shown under each image, supporting that our method makes the source image more joyful. A quantitative evaluation measuring the similarity of two probability distributions with 4 metrics M ($M \in \{KLD, BC, CD, EMD\}$) is shown on the right, where D_{M_s} is the distance between source and target distributions, and $D_{M_{tr}}$ is the distance between transformed and target distributions. For each metric, the better number is displayed in bold. Under all 4 measures, the transformed image evokes more similar emotions to the target image versus the source image.