



# Predicting Continuous Probability Distribution of Image Emotions in Valence-Arousal Space

Sicheng Zhao, Hongxun Yao, Xiaolei Jiang

School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China.  
 zsc@hit.edu.cn, h.yao@hit.edu.cn, xljiang@hit.edu.cn

## ABSTRACT

Previous works on image emotion analysis mainly focused on assigning a dominated emotion category or the average dimension values to an image for affective image classification and regression. However, this is often insufficient in many applications, as the emotions that are evoked in viewers by an image are highly subjective and different. In this paper, we propose to predict the continuous probability distribution of dimensional image emotions represented in valence-arousal space. By the statistical analysis on the constructed Image-Emotion-Social-Net dataset, we represent the emotion distribution as a Gaussian mixture model (GMM), which is estimated by the EM algorithm. Then we extract commonly used features of different levels for each image. Finally, we formulate the emotion distribution prediction as a multi-task shared sparse regression (MTSSR) problem, which is optimized by iteratively reweighted least squares. Besides, we introduce three baseline algorithms. Experiments conducted on the Image-Emotion-Social-Net dataset demonstrate the superiority of the proposed method, as compared to some state-of-the-art approaches.

## Categories and Subject Descriptors

H.3.1 [Information storage and retrieval]: Content Analysis and Indexing; J.5 [Computer Applications]: Arts and Humanities

## General Terms

Algorithms, Human Factors, Experimentation, Performance

## Keywords

Image Emotion Distribution; Valence-Arousal; Gaussian Mixture Model; Sparse Regression

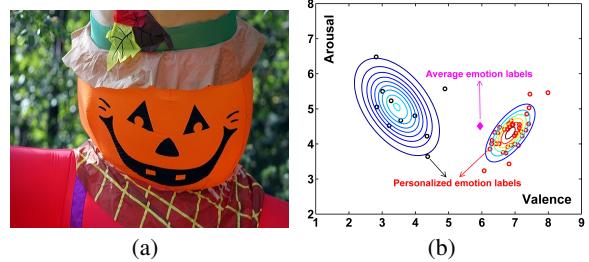
## 1. INTRODUCTION

Images can convey rich semantics and evoke strong emotions in viewers. Understanding the perceived emotions in images has been widely studied recently due to its vital importance in many applications of various domains, ranging from entertainment to education and advertisement [1, 13, 3]. However, this is a non-trivial

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

MM'15, October 26–30, 2015, Brisbane, Australia.

© 2015 ACM. ISBN 978-1-4503-3459-4/15/10 ...\$15.00.  
 DOI: <http://dx.doi.org/10.1145/2733373.2806354>

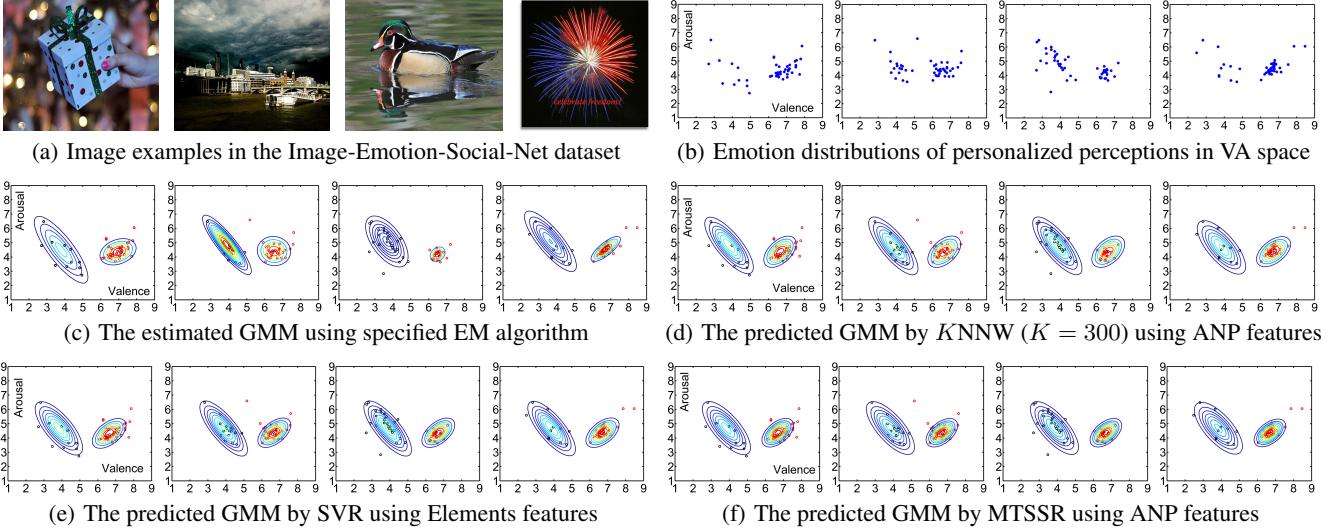


**Figure 1: The differences between affective image regression and emotion distribution prediction.** (a) is the original image, (b) illustrates the differences, where the hollow points are the perceived emotion labels in VA space, the magenta diamond point is the predicted average VA scores by traditional affective image regression, while the contour lines of GMM are the predicted emotion distribution by the proposed method.

problem due to the great challenges of affective gap and subjective evaluation [13]. According to the emotion representation models, including categorical emotion states (CES) and dimensional emotion space (DES) [13], related works on image emotion analysis can be classified into three different tasks: affective image classification [11, 7, 4, 6, 1, 13, 3], regression [6, 13] and retrieval [15].

Existing works mainly focused on finding features that can express emotions better to bridge the affective gap. Based on the extracted features, state-of-the-art methods tried to assign a dominated emotion category or the average dimension values to an image for affective image classification and regression with CES and DES models, respectively. Holistic features including Wiccest features and Gabor features were extracted to classify image emotions in [11]. Machajdik *et al.* [7] extracted features inspired from psychology and art theory, such as *color*, *texture* and *composition*. Lu *et al.* [6] investigated the computability of emotion through *shape* features. Zhao *et al.* [13] proposed to extract more interpretable emotion features based on principles-of-art, such as *balance*, *contrast* and *variety*, which were demonstrated to have stronger link to emotions than the elements-of-art. Visual sentiment ontology and detectors are proposed to detect high-level adjective noun pairs (ANPs) based on large-scale social multimedia data [1, 3]. Yuan *et al.* [12] used mid-level scene attributes for binary sentiment classification. Simple social correlation features are explored for emotion classification of social network images [4].

However, just predicting the dominated emotion is insufficient in many applications, as the emotions that are evoked in different viewers by an image are highly subjective and different [14], due to the influence of social and cultural backgrounds [5]. For example, in a test of image emotions to personal interest estimation,



**Figure 2: Image examples, related emotion labels and the predicted emotion distributions on the Image-Emotion-Social-Net dataset.**

the personalized emotion perceptions are required. In such cases, we need to take the subjective evaluation into account. Therefore, predicting the probability distribution of image emotions becomes more reasonable and useful in practice. To our knowledge, there have been few works tackling the subjective evaluation challenge and predicting the probability distribution of image emotions.

In this paper, we make an initial attempt to predict the probability distribution of image emotions in DES form, as shown in Figure 1. By the statistical analysis of personalized emotion perceptions on the constructed Image-Emotion-Social-Net dataset, we observe that the valence-arousal emotion labels can be well represented by a mixture of bidimensional Gaussian distributions. The expectation-maximization (EM) algorithm with specified initializations is used to estimate the parameters of GMM. We formulate the emotion distribution prediction as a multi-task shared sparse regression (MTSSR) problem, which is optimized by iteratively reweighted least squares. Experiments are conducted on the Image-Emotion-Social-Net dataset to demonstrate the effectiveness of the proposed emotion distribution prediction method.

## 2. PROBLEM DEFINITION

As there is no public dataset on image emotion distribution in VA space, we set up a large-scale dataset, named Image-Emotion-Social-Net dataset, with over 1 million images downloaded from Flickr. To get the personalized emotion labels, firstly we use traditional lexicon-based methods to obtain the text segmentation results of the title, tags and descriptions from uploaders for expected emotions and the comments from viewers for actual emotions. Then we compute the average value of valence, arousal and dominance of the segmentation results as ground truth for dimensional emotion representation based on recently published VAD norms of 13,915 English lemmas [9].

Some image examples and related personalized emotion labels are shown in Figure 2(a) and 2(b), from which we have the following observations: (1) The emotions evoked by an image in different viewers are truly subjective and different; Just assigning the average dimensional values of valence and arousal to an image is obviously not enough; (2) Though highly different, the perceived emotions follow certain distributions, which can be clearly grouped into two clusters, corresponding to the positive and negative senti-

ments; In each cluster, the VA emotion values are relatively stable; (3) The VA emotion labels can be well modeled by a mixture of two bidimensional Gaussian distributions.

Based on these observations, we define the distribution of VA emotion labels as a GMM by

$$p(\mathbf{x}; \boldsymbol{\theta}) = \sum_{l=1}^L \pi_l \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l), \quad (1)$$

where  $\mathbf{x} = (v, a)$  is pair-wise VA emotion labels,  $\boldsymbol{\mu}_l$  and  $\boldsymbol{\Sigma}_l$  are the mean vector and covariance matrix of the  $l$ th Gaussian component, while  $\pi_l$  is the mixing coefficient, which satisfies  $\pi_l \geq 0$  and  $\sum_{l=1}^L \pi_l = 1$ . In this paper, the number of Gaussian components is 2, i.e.  $L = 2$  and  $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \pi_1, \pi_2)$ . It should be noted that the number of Gaussian components  $L$  can be easily enlarged if more personalized emotion labels are obtained.

The EM algorithm is used to estimate the parameters of GMM. Specifically, the initializations are obtained by firstly partitioning the VA labels into two clusters based on whether valence is greater than 5 and then computing the mean vector  $\boldsymbol{\mu}_l$  and covariance matrix  $\boldsymbol{\Sigma}_l$  of each cluster. The mixing coefficients are set as the proportions of related VA labels in each cluster to the total labels. In experiment, the EM algorithm is converged in 6.28 steps on average. Some estimated results of GMM are shown in Figure 2(c).

Suppose we have  $N$  training images  $\mathbf{f}_1, \dots, \mathbf{f}_N$ , the emotion distributions are  $p_1(\mathbf{x}; \boldsymbol{\theta}_1), \dots, p_N(\mathbf{x}; \boldsymbol{\theta}_N)$ , where  $\boldsymbol{\theta}_n = (\boldsymbol{\mu}_{n1}, \boldsymbol{\Sigma}_{n1}, \boldsymbol{\mu}_{n2}, \boldsymbol{\Sigma}_{n2}, \pi_{n1}, \pi_{n2})$  are the parameters of the  $n$ th emotion distribution ( $n = 1, \dots, N$ ). Similarly, suppose we have  $M$  test images  $\mathbf{g}_1, \dots, \mathbf{g}_M$  with ground truth emotion distributions  $q_1(\mathbf{x}; \boldsymbol{\vartheta}_1), \dots, q_M(\mathbf{x}; \boldsymbol{\vartheta}_M)$ , where  $\boldsymbol{\vartheta}_m (m = 1, \dots, M)$  are the distribution parameters. Then our goal is to predict the emotion distribution parameters  $\hat{\boldsymbol{\vartheta}}_m$  based on  $\{\mathbf{f}_n, \boldsymbol{\theta}_n\}_{n=1}^N$  for each  $\mathbf{g}_m$ . That is

$$f : (\{\mathbf{f}_n, \boldsymbol{\theta}_n\}_{n=1}^N, \mathbf{g}_m) \rightarrow \hat{\boldsymbol{\vartheta}}_m. \quad (2)$$

## 3. EMOTION DISTRIBUTION PREDICTION ALGORITHMS

The emotion distribution prediction task of Equ. (2) can be viewed as a regression problem. In this section, we briefly introduce the extracted emotion features and detail the proposed MTSSR method together with several baseline algorithms.

As shown in [15], there are various types of features that may contribute to the perceptions of image emotions. Similar to [15], we extract commonly used emotion features of different levels and generalities for each image, including low-level GIST and elements-of-art [7], mid-level attributes [8] and principles-of-art [13], and high-level ANPs [1] and expressions [10]. The extracted features are abbreviated as GIST, Elements, Attributes, Principles, ANP and Expressions with dimension 512, 48, 102, 165, 1200 and 8, respectively. Please refer to [15] for details.

### 3.1 K-Nearest Neighbor Weighting

The emotion distribution parameters  $\theta_n$  ( $n = 1, \dots, N$ ) of all training images are considered as basis functions.  $K$ -nearest neighbor weighting ( $KNNW$ ) selects the top  $K$  most similar training images and estimates  $\hat{\vartheta}_m$  by weighting the selected basis functions

$$\hat{\vartheta}_m = \frac{\sum_{k=1}^K s_{t_k} \theta_{t_k}}{\sum_{k=1}^K s_{t_k}}, \quad (3)$$

where  $s_n = \exp(-\frac{d(\mathbf{g}_m, \mathbf{f}_n)}{\sigma})$  is the similarity between images  $\mathbf{g}_m$  and  $\mathbf{f}_n$ ,  $d(\cdot, \cdot)$  is a specified distance function,  $\sigma$  is set as the average distance of all the training images, while  $s_{t_1}, \dots, s_{t_K}$  are the top  $K$  largest similarities. In experiment, the Euclidean distance is used for  $d(\cdot, \cdot)$  and each  $\theta_n$ ,  $\hat{\vartheta}_m$  is reshaped as a column vector for convenience.

### 3.2 Support Vector Regression

Support vector regression (SVR) aims to find support vectors which lie on the maximum margin hyperplanes in feature space and contribute to predictions. Training SVR means solving

$$\min \frac{1}{2} \|\mathbf{w}_i\|_2, \text{s.t. } \begin{cases} \theta_{ni} - \langle \mathbf{w}_i, \mathbf{f}_n \rangle - b \leq \epsilon, \\ \langle \mathbf{w}_i, \mathbf{f}_n \rangle + b - \theta_{ni} \leq \epsilon, \end{cases} \quad (4)$$

where the target value  $\theta_{ni}$  is the  $i$ th component of  $\theta_n$  ( $n = 1, \dots, N$ ), the inner product plus intercept  $\langle \mathbf{w}_i, \mathbf{f}_n \rangle + b$  is the prediction for that sample, and  $\epsilon$  is a free parameter that serves as a threshold. After optimization, we can predict  $\hat{\vartheta}_m$  by  $\hat{\vartheta}_m = \langle \mathbf{w}_i, \mathbf{g}_m \rangle + b$ . We use the LIBSVM<sup>1</sup> toolbox with linear kernel (for fast speed) to implement SVR for emotion distribution prediction.

### 3.3 Shared Sparse Regression

Suppose  $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_N]$ ,  $\Theta = [\theta_1, \dots, \theta_N]$ . The basic idea of shared sparse regression (SSR) is that  $\mathbf{g}_m$  and  $\hat{\vartheta}_m$  can be written in terms of bases  $\mathbf{F}$  and  $\Theta$  respectively, but with shared sparse coefficients  $\phi_m$ . That is

$$\mathbf{g}_m = \mathbf{F}\phi_m \quad \text{and} \quad \hat{\vartheta}_m = \Theta\phi_m, \quad (5)$$

where  $\phi_m$  is obtained by

$$\phi_m^* = \underset{\phi_m}{\operatorname{argmin}} \|\mathbf{F}\phi_m - \mathbf{g}_m\|^2 + \eta \|\phi_m\|_1, \quad (6)$$

where  $\eta$  is a regularization coefficient. Equ. (6) can be solved using iteratively reweighted least squares [2] (similar to Section 3.4).

### 3.4 Multi-Task Shared Sparse Regression

The above three baselines either model one test image or predict one target value each time, without jointly combining them together to explore the latent correlation. Multi-task shared sparse regression (MTSSR) utilizes this information. Suppose  $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_M]$ ,

$\Omega = [\hat{\vartheta}_1, \dots, \hat{\vartheta}_M]$ , MTSSR jointly predicts  $\Omega$  for  $\mathbf{G}$  by letting the test features and the target values share the same coefficients  $\Phi$  on training data  $\mathbf{F}$  and  $\Theta$  as follows

$$\mathbf{G} = \mathbf{F}\Phi \quad \text{and} \quad \Omega = \Theta\Phi. \quad (7)$$

$\Phi \in \mathbb{R}^{N \times M}$  is obtained by solving the following convex optimization problem

$$\min_{\Phi} \|\mathbf{F}\Phi - \mathbf{G}\|^2 + \eta_1 \|\Phi\|_1 + \eta_2 \|\Phi\|_{2,1}, \quad (8)$$

where  $\eta_1$  and  $\eta_2$  are regularization coefficients,  $\|\cdot\|_{2,1}$  denotes the

$$\ell_{2,1}\text{-norm of a matrix } \|\Phi\|_{2,1} = \sum_{n=1}^N \sqrt{\sum_{m=1}^M \phi_{n,m}^2}.$$

We employ the iteratively reweighted least squares [2] to optimize Equ. (8). The components of Equ. (8) are transformed by

$$\|\Phi\|_1 = \sum_{n,m} |\phi_{n,m}| \simeq \sum_{n,m} \frac{\phi_{n,m}^2}{|\phi_{n,m}| + \varepsilon}, \quad (9)$$

$$\|\Phi\|_{2,1} = \sum_n \sqrt{\sum_m \phi_{n,m}^2} \simeq \sum_n \left( \frac{\sum_m \phi_{n,m}^2}{\sqrt{\sum_m \phi_{n,m}^2} + \varepsilon} \right). \quad (10)$$

Let  $\varphi_{n,m} = \frac{1}{|\phi_{n,m}| + \varepsilon}$  and  $\psi_n = \frac{1}{\sqrt{\sum_m \phi_{n,m}^2} + \varepsilon}$ , then the objective function of Equ.(8) is transformed to

$$\begin{aligned} \mathcal{J}(\Phi) &\simeq \sum_m \|\mathbf{F}\phi_m - \mathbf{g}_m\|^2 + \sum_{n,m} (\eta_1 \varphi_{n,m} + \eta_2 \psi_n) \phi_{n,m}^2 \\ &= \sum_m \|\mathbf{F}\phi_m - \mathbf{g}_m\|^2 + \sum_m \phi_m^\top \mathbf{W}_m \phi_m, \end{aligned} \quad (11)$$

where  $\mathbf{W}_m$  is a diagonal matrix with  $\mathbf{W}_m(n,n) = \eta_1 \varphi_{n,m} + \eta_2 \psi_n$ .  $\min \mathcal{J}(\Phi)$  is a quadratic programming problem, which can be easily solved by off-the-shelf optimization methods. This process is iteratively updated until convergence.

## 4 EXPERIMENTS

To evaluate the effectiveness of the proposed method for continuous distribution prediction of dimensional image emotions, we carried out experiments on the Image-Emotion-Social-Net dataset (see Section 2). Totally we selected 18,700 images with more than 20 VA labels each for experiment. The data can be downloaded at <https://sites.google.com/site/schzhao/>.

### 4.1 Evaluation Criteria

The Kullback-Leibler divergence and the log likelihood (abbreviated as  $KL$  and  $LLH$ ) are used as the evaluation metric. The KL divergence of the predicted distribution  $\hat{q}_m(\mathbf{x}; \hat{\vartheta}_m)$  from the ground truth distribution  $q_m(\mathbf{x}; \vartheta_m)$  is defined as

$$KL(q_m || \hat{q}_m) = - \int q_m(\mathbf{x}; \vartheta_m) \ln \left\{ \frac{\hat{q}_m(\mathbf{x}; \hat{\vartheta}_m)}{q_m(\mathbf{x}; \vartheta_m)} \right\} d\mathbf{x}. \quad (12)$$

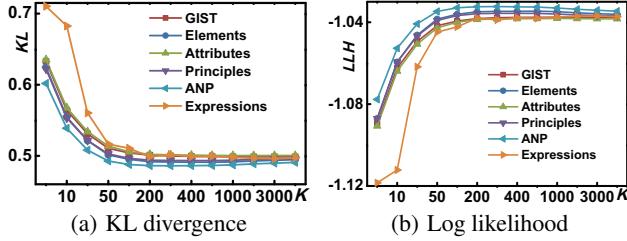
In practice,  $KL(q_m || \hat{q}_m)$  is approximated by a finite sum of the points  $\{\mathbf{s}_1, \dots, \mathbf{s}_S\}$  sampled following distribution  $q_m(\mathbf{x}; \vartheta_m)$  by

$$KL(q_m || \hat{q}_m) \simeq \frac{1}{S} \sum_{n=1}^S \left\{ \ln q_m(\mathbf{s}_n; \vartheta_m) - \ln \hat{q}_m(\mathbf{s}_n; \hat{\vartheta}_m) \right\}. \quad (13)$$

The log likelihood metric is computed based on the actual VA labels  $\{\mathbf{x}_{m1}, \dots, \mathbf{x}_{mR_m}\}$  by

$$LLH(m) = \frac{1}{R_m} \log \prod_{r=1}^{R_m} \hat{q}_m(\mathbf{x}_{mr}; \hat{\vartheta}_m) = \frac{1}{R_m} \sum_{r=1}^{R_m} \log \hat{q}_m(\mathbf{x}_{mr}; \hat{\vartheta}_m). \quad (14)$$

<sup>1</sup><http://www.csie.ntu.edu.tw/~cjlin/libsvm/>



**Figure 3: The influence of  $K$  in KNNW on continuous emotion distribution prediction using different features on measurement (a) KL divergence and (b) Log likelihood.**

$KL \geq 0$  and lower value indicates better performance, while  $LLH \leq 0$  and higher value represents better performance. In experiment, the data was separated into a training set and a test set using K-fold Cross Validation ( $K=5$ ). We computed the average  $KL$  and  $LLH$  together with the standard deviation of the  $K$  folds.

## 4.2 Results and Discussions

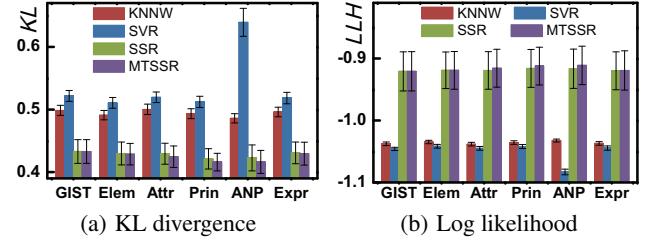
Firstly, we investigated the influence of  $K$  in KNNW on emotion distribution prediction ( $K = 5, 10, 20, 50, 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000$ ). The results are illustrated in Figure 3. For clarity, the standard deviations are not shown. It is clear to see that the best  $K$  is dependent on the extracted features and is almost consistent on  $KL$  and  $LLH$  for each feature.  $K = 4000, 500, 1000, 400, 300$  and  $3000$  perform best for features GIST, Elements, Attributes, Principles, ANP and Expressions, respectively. These  $K$ s are selected as baselines for comparison with MTSSR.

The regularization parameters  $\eta$  and  $\eta_1, \eta_2$  can influence the performance of SSR and the proposed MTSSR. In experiment, they are selected by grid search. Due to page limit, we do not show the detailed influences here. Using the best  $\eta$  and  $\eta_1, \eta_2$  for each kind of features, we compared the performance of the proposed method with the three baselines on different features. The results measured by  $KL$  and  $LLH$  are illustrated in Figure 4. Some detailed prediction results of different methods using related best features are shown in Figure 2(d), 2(e) and 2(f).

From these results, we can find that (1)  $KL$  and  $LLH$  are dependent on both the features and the models; they are relatively consistent to measure the performance of distribution prediction; (2) For almost all the features, the proposed MTSSR model performs the best, which benefit from the exploration of latent information between different tasks; (3) The best features are ANP, Elements, Principles and ANP for KNNR, SVR, SSR and MTSSR, respectively; Generally, the low-level generic features perform the worst, which indicates that they cannot represent image emotions well because of the largest "affective gap"; Mid-level and the high-level features have stronger link to image emotions, which is consistent with the conclusions in [15]; (4) Though simple, the KNNW outperforms SVR on average in emotion distribution prediction.

## 5. CONCLUSION

In this paper, we proposed to predict the continuous probability distribution of image emotions represented in VA space, which can be viewed as an initial attempt to measure the subjective evaluation of human perceptions. We presented multi-task shared sparse learning as the learning model and optimized it by iteratively reweighted least squares. Besides, we provided another three baseline algorithms. Experiments conducted on the Image-Emotion-Social-Net dataset demonstrated the effectiveness of the proposed method.



**Figure 4: Performance comparison between the proposed method and the three baselines on emotion distribution prediction using different features measured by  $KL$  and  $LLH$ .**

The predicted emotion distribution can be explored in many applications, such as affective image retrieval and emotion transfer. For further studies, we will consider exploring social related factors for emotion distribution prediction. Consistently combining and fusing multi-modal features in MTSSR may further improve the prediction performance, which is also worth studying.

## 6. ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (No. 61472103) and Key Program (No. 61133003).

## 7. REFERENCES

- [1] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *ACM International Conference on Multimedia*, pages 223–232, 2013.
- [2] C. Chen, J. Huang, L. He, and H. Li. Preconditioning for accelerated iteratively reweighted least squares in structured sparsity reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2713–2720, 2014.
- [3] T. Chen, F. X. Yu, J. Chen, Y. Cui, Y.-Y. Chen, and S.-F. Chang. Object-based visual sentiment concept analysis and application. In *ACM International Conference on Multimedia*, pages 367–376, 2014.
- [4] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. Can we understand van gogh's mood? learning to infer affects from images in social networks. In *ACM International Conference on Multimedia*, pages 857–860, 2012.
- [5] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, and J. Luo. Aesthetics and emotions in images. *IEEE Signal Processing Magazine*, 28(5):94–115, 2011.
- [6] X. Lu, P. Suryanarayanan, R. B. Adams Jr, J. Li, M. G. Newman, and J. Z. Wang. On shape and the computability of emotions. In *ACM International Conference on Multimedia*, pages 229–238, 2012.
- [7] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM International Conference on Multimedia*, pages 83–92, 2010.
- [8] G. Patterson and J. Hays. Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2751–2758, 2012.
- [9] A. B. Warriner, V. Kuperman, and M. Brysbaert. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207, 2013.
- [10] P. Yang, Q. Liu, and D. N. Metaxas. Exploring facial expressions with compositional features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2638–2644, 2010.
- [11] V. Yanulevskaya, J. Van Gemert, K. Roth, A.-K. Herbold, N. Sebe, and J.-M. Geusebroek. Emotional valence categorization using holistic image features. In *IEEE International Conference on Image Processing*, pages 101–104, 2008.
- [12] J. Yuan, S. McDonough, Q. You, and J. Luo. Sentribute: image sentiment analysis from a mid-level perspective. In *ACM International Workshop on Issues of Sentiment Discovery and Opinion Mining*, page 10, 2013.
- [13] S. Zhao, Y. Gao, X. Jiang, H. Yao, T.-S. Chua, and X. Sun. Exploring principles-of-art features for image emotion recognition. In *ACM International Conference on Multimedia*, pages 47–56, 2014.
- [14] S. Zhao, H. Yao, X. Jiang, and X. Sun. Predicting discrete probability distribution of image emotions. In *IEEE International Conference on Image Processing*, 2015.
- [15] S. Zhao, H. Yao, Y. Yang, and Y. Zhang. Affective image retrieval via multi-graph learning. In *ACM International Conference on Multimedia*, pages 1025–1028, 2014.