Understanding Emotional Impact of Images Using Bayesian Multiple Kernel Learning

He Zhang*, Mehmet Gönen, Zhirong Yang, Erkki Oja

Department of Information and Computer Science, Aalto University School of Science, FI-00076 Aalto, Espoo, Finland

Abstract

Affective classification and retrieval of multimedia such as audio, image, and video have become emerging research areas in recent years. The previous research focused on designing features and developing feature extraction methods. Generally, a multimedia content can be represented with different feature representations (i.e., views). However, the most suitable feature representation related to people's emotions is usually not known a priori. We propose here a novel Bayesian multiple kernel learning algorithm for affective classification and retrieval tasks. The proposed method can make use of different representations simultaneously (i.e., multiview learning) to obtain a better prediction performance than using a single feature representation (i.e., single-view learning) or a subset of features, with the advantage of automatic feature selections. In particular, our algorithm has been implemented within a multilabel setup to capture the correlation between emotions, and the Bayesian formulation enables our method to produce probabilistic outputs for measuring a set of emotions triggered by a single image. As a case study, we perform classification and retrieval experiments with our algorithm for predicting people's emotional states evoked by images, using generic low-level image features. The empirical results with our approach on the widely-used `International Affective Picture System (IAPS)` data set outperforms several existing methods in terms of classification

*Corresponding author.

Email addresses: `he.zhang@aalto.fi` (He Zhang), `mehmet.gonen@aalto.fi` (Mehmet Gönen), `zhirong.yang@aalto.fi` (Zhirong Yang), `erkki.oja@aalto.fi` (Erkki Oja)

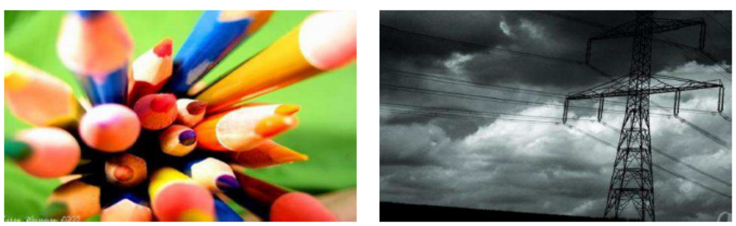
Preprint submitted to Neurocomputing *August 10, 2014*

performance and results interpretability.

Keywords: Image emotions, multiple kernel learning, multiview learning, variational approximation, low-level image features

1. Introduction

Affective computing [1] aims to help people communicate, understand, and respond better to affective information such as audio, image, and video in a way that takes into account the user's emotional states. Among the emotional stimuli, affective image classification and retrieval has attracted increasing research attention in recent years, due to the rapid expansion of the digital visual libraries on the Web. While most of the current Content-Based Image Retrieval (CBIR) systems [2] are designed for recognizing objects and scenes such as plants, animals, outdoor places etc., an Emotional Semantic Image Retrieval (ESIR) system [3] aims at incorporating the user's affective states to enable queries like "beautiful flowers", "cute dogs", "exciting games", etc.



(a) Amusement (b) Fear

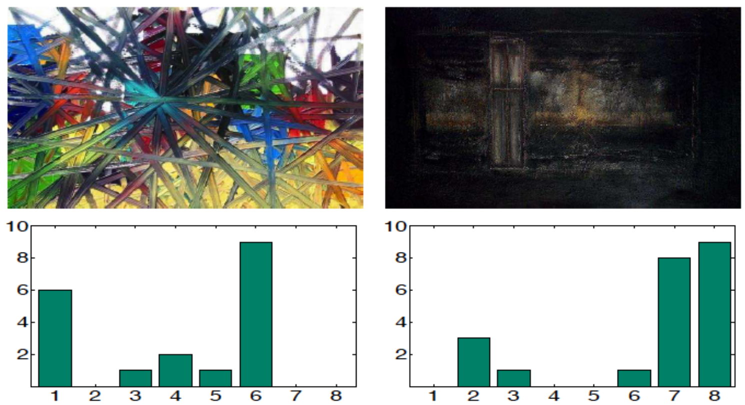
Figure 1: Example images from a photo sharing site (`ArtPhoto` [4]) with the ground truth labels of Amusement and Fear.

Though emotions are highly subjective human factors, still they have stability and generality across different people and cultures [5]. As an example, Figure 1 shows two pictures taken from a photo sharing site (`ArtPhoto` [4]). The class labels of "Amusement" and "Fear" are determined by the emotion that has received the most votes from people. Intuitively, an "Amusement" picture usually makes people feel pleasant or induces high valence, whereas a "Fear" picture may induce low valence but high arousal to the viewer.

In analogy to the concept of "semantic gap" that implies the limitations of image recognition techniques, the "affective gap" can be defined as "the lack of coincidence between the measurable signal properties, commonly referred to as features, and the expected affective state in which the user is brought by perceiving the signal" [6]. Concerning the studies related to image affect recognition, three major challenges can be identified: (a) the modeling of affect, (b) the extraction of image features to reflect affective states, and (c) the building of classifiers to bridge the "affective gap".

Most of the current works (e.g., [5, 7, 4, 8]) use descriptive words (e.g., the scenario in Figure 1) to represent affective space. To obtain the ground truth label for learning, each image is assigned with a single emotional label among various emotional categories based on the maximum votes from the viewers. However, an image can usually evoke a mixture of affective feelings in people rather than a single one. Furthermore, the emotions often conceptually correlate with each other in the affective space. For example, the two paintings shown in Figure 2 are labeled as "Excitement" and "Sad(ness)" respectively according to the maximum votes (from the web survey in [4]). Nevertheless, by examining the votes from the viewers, each image actually has evoked a distribution of emotions rather than a single one. Moreover, the correlations can be observed between certain emotions. For example, "Amusement" is closely associated with "Excitement", and "Fear" often comes with "Sadness".

Feature extraction is a prerequisite step for image classification and retrieval tasks [2], especially for the recognition of emotions induced by pictures or artworks. In the literature, much effort has been spent on designing features specific to image affect recognition (e.g., [9, 7, 4, 10, 11]). Other works (e.g., [12, 13, 14, 8]) used the generic low-level color, shape, and texture features for detecting the image emotions. Concerning the inference, supervised learning has been used more often than unsupervised learning for inferring the image emotions. Among the classifiers, Support Vector Machines (SVMs) have been adopted by most of the works (e.g., [13, 15, 7, 16, 8]). Since the most suitable feature representation or subset related to people's emotions is not known a



(a) Excitement (b) Sad

Figure 2: Example images from an online user survey showing that images can evoke mixed feelings in people instead of a single one [4]. The x-axis shows emotions (from left to right): Amusement, Anger, Awe, Contentment, Disgust, Excitement, Fear, Sad. The y-axis shows the number of votes.

priori, feature selection has to be done for better prediction performance prior to the final prediction, which increases the computational complexity. Instead of using a single representation or view, we can also make use of different representations or views at the same time. This implies that multiview learning [17] is preferred to single-view learning. Multiview learning with kernel-based methods belongs to the framework of Multiple Kernel Learning (MKL), which is a principled way of combining kernels calculated on different views to obtain a better prediction performance than single-view learning methods (see [18] for a recent survey).

In this paper, we propose a novel Bayesian multiple kernel learning algorithm for affective classification and retrieval tasks with multiple outputs and feature representations. Thanks to the MKL framework, our method can learn the feature representation weights by itself according to the data and task at hand without an explicit feature selection step, which makes the interpretation easy and straightforward. Our method has been implemented within a multilabel setup in order to capture the correlations between emotions. Due to its probabilistic nature, our method is able to produce probabilistic values for measuring the intensities of a set of emotions triggered by a single image. As a case study, we conduct classification and retrieval experiments with our proposed approach for predicting people's emotional states evoked by images, using conventional low-level color, shape, and texture image features. The experimental results on the widely-used `International Affective Picture System (IAPS)` data set show that our proposed Bayesian MKL approach outperforms other existing methods in terms of classification performance, feature selection capacity, and results interpretability.

Our contributions are thus two-fold:

1. Instead of single view representation, a multiview learning with kernel-based method has been applied to emotional image recognition, with the advantages of better prediction performance, automatic feature selection, and interpretation of image emotional impact.
2. A novel Bayesian multiple kernel learning algorithm with multiple outputs and feature representations has been proposed for affective classification and retrieval tasks. Our method is able to capture the correlations between emotions and give probabilistic outputs for measuring the intensities of a distribution of emotions triggered by an image.

We start in the following section with a concise review on the related work. Section 3 gives the mathematical details of the proposed method. In Section 4, the experimental results on affective image classification and retrieval are reported. Finally, the conclusions and future work are presented in Section 5.

2. Related Work

In this section, we review the works related to image affect recognition, with an emphasis on affective modeling, feature extraction, and classifier construction.

Affect has been conceptualized in psychology [19]. There are two primary ways to modeling affect: the dimensional approach and the discrete approach. The dimensional approach [20] describes affect within a 3D continuous space along *Valence*, *Arousal*, and *Dominance*. Valence is typically characterized as the affective states ranging from pleasant or positive to unpleasant or negative. Arousal is characterized as the state of being awake or reactive to stimuli, ranging from calm to excited. Dominance denotes power and influence over others, ranging from no control to full control. The discrete approach describes affect with a list of descriptive or adjective words (as the example given in Figure 2). A popular example is Ekman's six basic emotion categories, namely, *happiness*, *sadness*, *fear*, *anger*, *disgust*, and *surprise*. Most of the current works [12, 5, 7, 4, 10, 8] related to image affect recognition focus on recognizing the discrete emotions extended from these basic emotions. For example, positive emotions may include *amusement*, *awe*, *contentment*, and *excitement*, while the negative emotions consist of *anger*, *disgust*, *fear*, and *sadness* [21]. For our work, we adopt the discrete approach as well.

Features specific to affective image classification have been developed in [9, 7, 4, 10, 11]. For example, the authors in [9] and [4] designed color features based on Itten's contrast theory. Specifically, the authors in [9] exploited semiotic principles to represent the visual content at the expressive level, while the authors in [4] used the composition features such as the low depth-of-field indicators, rule of thirds, and proportion of face and skin pixels in images, which have been found useful for aesthetics. The luminance-warm-cool and saturation-warm-cool color histograms were derived in [7] based on the fuzzy theory. In [10], the authors investigated the relationship between shape and emotions. They found that roundness and complexity of shapes are fundamental to understanding emotions. On the contrary, the conventional low-level image features have been adopted in [12, 13, 14, 8]. For example, a large set of generic color, shape, and texture image features have been used in [14] and [8]. These low-level features were extracted from both the raw images and compound image transforms such as color transform and edge transform, which were found highly effective earlier in face recognition and the classification of painters and schools of art. In our work, we also use the conventional low-level image features, and we show later in the experiments that the proposed method can learn well enough to predict image emotions by using low-level features.

As for classifiers, SVM [22] is the most favorite one and has been used in [13, 15, 7, 16, 8]. Others include the naive Bayes classifier [23] used in [11, 4, 10] and the regression trees [24] used in [15]. In this paper, we follow the Bayesian approach. As a methodological contribution, the proposed algorithm is the first multiple kernel learning algorithm that combines multiview learning and multilabel learning with full Bayesian treatment. There are existing Bayesian MKL algorithms and multilabel learning methods applied to image classification problems (e.g., [25]) but there is no previous study on a coupled approach. In this case, our method has the advantage of utilizing the emotional correlations in image affect recognition.

3. Proposed Method

In order to benefit from the correlation between the class labels in a multilabel learning scenario, we assume a common set of kernel weights and perform classification for all labels with these weights but using a distinct set of classification parameters for each label. This approach can also be interpreted as using a common similarity measure by sharing the kernel weights between the labels.

The notation we use throughout the manuscript is given in Table 1. The superscripts index the rows of matrices, whereas the subscripts index the columns of matrices and the entries of vectors. $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the normal distribution with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. $\mathcal{G}(\cdot; \alpha, \beta)$ denotes the gamma distribution with the shape parameter α and the scale parameter β. $\delta(\cdot)$ denotes the Kronecker delta function that returns 1 if its argument is true and 0 otherwise.

Figure 3 illustrates the proposed probabilistic model for multilabel binary classification with a graphical model. We extended the model presented in [26] by trying to capture the correlation between the class labels with the help of shared kernel weights. The kernel matrices $\{\mathbf{K}_1, \ldots, \mathbf{K}_P\}$ are used to calculate

Table 1: List of notation.

N	Number of training instances
P	Number of kernels
L	Number of output labels
$\{\mathbf{K}_1, \ldots, \mathbf{K}_P\} \in \mathbb{R}^{N \times N}$	Kernel matrices
$\mathbf{A} \in \mathbb{R}^{N \times L}$	Weight matrix
$\mathbf{\Lambda} \in \mathbb{R}^{N \times L}$	Priors for weight matrix
$\{\mathbf{G}_1, \ldots, \mathbf{G}_L\} \in \mathbb{R}^{P \times N}$	Intermediate outputs
$\boldsymbol{e} \in \mathbb{R}^P$	Kernel weight vector
$\boldsymbol{\omega} \in \mathbb{R}^P$	Priors for kernel weight vector
$\boldsymbol{b} \in \mathbb{R}^L$	Bias vector
$\boldsymbol{\gamma} \in \mathbb{R}^L$	Priors for bias vector
$\mathbf{F} \in \mathbb{R}^{L \times N}$	Auxiliary matrix
$\mathbf{Y} \in \{\pm 1\}^{L \times N}$	Label matrix

intermediate outputs using the weight matrix $\mathbf{A}$. The intermediate outputs $\{\mathbf{G}_1, \ldots, \mathbf{G}_L\}$, kernel weights $\boldsymbol{e}$, and bias parameters $\boldsymbol{b}$ are used to calculate the classification scores. Finally, the given class labels $\mathbf{Y}$ are generated from the auxiliary matrix $\mathbf{F}$, which is introduced to make the inference procedures efficient [27]. We formulated a variational approximation procedure for inference in order to have a computationally efficient algorithm.

The distributional assumptions of our proposed model are defined as

$$\begin{aligned}
\lambda_o^i &\sim \mathcal{G}(\lambda_o^i; \alpha_\lambda, \beta_\lambda) & \forall (i, o) \\
a_o^i | \lambda_o^i &\sim \mathcal{N}(a_o^i; 0, (\lambda_o^i)^{-1}) & \forall (i, o) \\
g_{o,i}^m | \boldsymbol{a}_o, \boldsymbol{k}_{m,i} &\sim \mathcal{N}(g_{o,i}^m; \boldsymbol{a}_o^\top \boldsymbol{k}_{m,i}, 1) & \forall (o, m, i) \\
\gamma_o &\sim \mathcal{G}(\gamma_o; \alpha_\gamma, \beta_\gamma) & \forall o \\
b_o | \gamma_o &\sim \mathcal{N}(b_o; 0, \gamma_o^{-1}) & \forall o
\end{aligned}$$

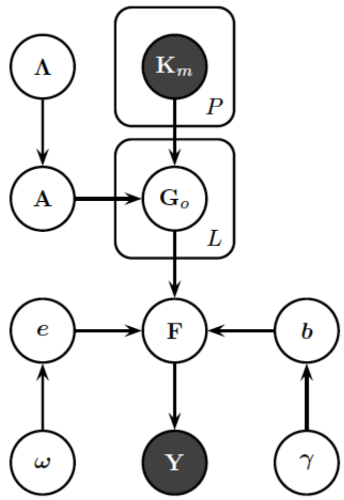


Figure 3: Graphical model for Bayesian multilabel multiple kernel learning.

$$
\begin{aligned}
\omega_m &\sim \mathcal{G}(\omega_m; \alpha_\omega, \beta_\omega) && \forall m \\
e_m | \omega_m &\sim \mathcal{N}(e_m; 0, \omega_m^{-1}) && \forall m \\
f_i^o | b_o, \boldsymbol{e}, \boldsymbol{g}_{o,i} &\sim \mathcal{N}(f_i^o; \boldsymbol{e}^\top \boldsymbol{g}_{o,i} + b_o, 1) && \forall (o, i) \\
y_i^o | f_i^o &\sim \delta(f_i^o y_i^o > \nu) && \forall (o, i)
\end{aligned}
$$

where the margin parameter ν is introduced to resolve the scaling ambiguity issue and to place a low-density region between two classes, similar to the margin idea in SVMs, which is generally used for semi-supervised learning [28]. As short-hand notations, all priors in the model are denoted by $\boldsymbol{\Xi} = \{\boldsymbol{\gamma}, \boldsymbol{\Lambda}, \boldsymbol{\omega}\}$, where the remaining variables by $\boldsymbol{\Theta} = \{\mathbf{A}, \boldsymbol{b}, \boldsymbol{e}, \mathbf{F}, \mathbf{G}_1, \ldots, \mathbf{G}_L\}$ and the hyper-parameters by $\boldsymbol{\zeta} = \{\alpha_\gamma, \beta_\gamma, \alpha_\lambda, \beta_\lambda, \alpha_\omega, \beta_\omega\}$. Dependence on $\boldsymbol{\zeta}$ is omitted for clarity throughout the manuscript.

The variational methods use a lower bound on the marginal likelihood using an ensemble of factored posteriors to find the joint parameter distribution [29]. Assuming independence between the approximate posteriors in the factorable ensemble can be justified because there is not a strong coupling between our model parameters. We can write the factorable ensemble approximation of the required posterior as

$$p(\mathbf{\Theta}, \mathbf{\Xi} | \{\mathbf{K}_m\}_{m=1}^P, \mathbf{Y}) \approx q(\mathbf{\Theta}, \mathbf{\Xi}) = \\ q(\mathbf{\Lambda}) q(\mathbf{A}) q(\mathbf{Z}) q(\{\mathbf{G}_o\}_{o=1}^L) q(\boldsymbol{\gamma}) q(\boldsymbol{\omega}) q(\boldsymbol{b}, \boldsymbol{e}) q(\mathbf{F})$$

and define each factor in the ensemble just like its full conditional distribution:

$$\begin{aligned}
q(\mathbf{\Lambda}) &= \prod_{i=1}^N \prod_{o=1}^L \mathcal{G}(\lambda_o^i; \alpha(\lambda_o^i), \beta(\lambda_o^i)) \\
q(\mathbf{A}) &= \prod_{o=1}^L \mathcal{N}(\boldsymbol{a}_o; \mu(\boldsymbol{a}_o), \Sigma(\boldsymbol{a}_o)) \\
q(\{\mathbf{G}_o\}_{o=1}^L) &= \prod_{o=1}^L \prod_{i=1}^N \mathcal{N}(\boldsymbol{g}_{o,i}; \mu(\boldsymbol{g}_{o,i}), \Sigma(\boldsymbol{g}_{o,i})) \\
q(\boldsymbol{\gamma}) &= \prod_{o=1}^L \mathcal{G}(\gamma_o; \alpha(\gamma_o), \beta(\gamma_o)) \\
q(\boldsymbol{\omega}) &= \prod_{m=1}^P \mathcal{G}(\omega_m; \alpha(\omega_m), \beta(\omega_m)) \\
q(\boldsymbol{b}, \boldsymbol{e}) &= \mathcal{N}\left(\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{e} \end{bmatrix}; \mu(\boldsymbol{b}, \boldsymbol{e}), \Sigma(\boldsymbol{b}, \boldsymbol{e}) \right) \\
q(\mathbf{F}) &= \prod_{o=1}^L \prod_{i=1}^N \mathcal{TN}(f_i^o; \mu(f_i^o), \Sigma(f_i^o), \rho(f_i^o))
\end{aligned}$$

where $\alpha(\cdot)$, $\beta(\cdot)$, $\mu(\cdot)$, and $\Sigma(\cdot)$ denote the shape parameter, the scale parameter, the mean vector, and the covariance matrix for their arguments, respectively. $\mathcal{TN}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma}, \rho(\cdot))$ denotes the truncated normal distribution with the mean vector $\boldsymbol{\mu}$, the covariance matrix $\mathbf{\Sigma}$, and the truncation rule $\rho(\cdot)$ such that $\mathcal{TN}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma}, \rho(\cdot)) \propto \mathcal{N}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma})$ if $\rho(\cdot)$ is true and $\mathcal{TN}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma}, \rho(\cdot)) = 0$ otherwise.

We can bound the marginal likelihood using Jensen's inequality:

$$\log p(\mathbf{Y} | \{\mathbf{K}_m\}_{m=1}^P) \geq \\ \mathrm{E}_{q(\mathbf{\Theta}, \mathbf{\Xi})}[\log p(\mathbf{Y}, \mathbf{\Theta}, \mathbf{\Xi} | \{\mathbf{K}_m\}_{m=1}^P)] - \mathrm{E}_{q(\mathbf{\Theta}, \mathbf{\Xi})}[\log q(\mathbf{\Theta}, \mathbf{\Xi})]$$

and optimize this bound by optimizing with respect to each factor separately until convergence. The approximate posterior distribution of a specific factor $\boldsymbol{\tau}$

can be found as

$$q(\boldsymbol{\tau}) \propto \exp\bigl(\mathrm{E}_{q(\{\boldsymbol{\Theta},\boldsymbol{\Xi}\}\setminus\boldsymbol{\tau})}[\log p(\mathbf{Y}, \boldsymbol{\Theta}, \boldsymbol{\Xi}|\{\mathbf{K}_m\}_{m=1}^{P})]\bigr).$$

For our model, thanks to the conjugacy, the resulting approximate posterior distribution of each factor follows the same distribution as the corresponding factor.

3.1. Inference Details

The approximate posterior distribution of the priors of the precisions for the weight matrix can be found as a product of gamma distributions:

$$q(\boldsymbol{\Lambda}) = \prod_{i=1}^{N}\prod_{o=1}^{L}\mathcal{G}\left(\lambda_o^i; \alpha_\lambda + \frac{1}{2}, \left(\frac{1}{\beta_\lambda} + \frac{\widetilde{(a_o^i)^2}}{2}\right)^{-1}\right) \tag{1}$$

where the tilde notation denotes the posterior expectations as usual, i.e., $\widetilde{h(\boldsymbol{\tau})} = \mathrm{E}_{q(\boldsymbol{\tau})}[h(\boldsymbol{\tau})]$. The approximate posterior distribution of the weight matrix is a product of multivariate normal distributions:

$$q(\mathbf{A}) = \prod_{o=1}^{L}\mathcal{N}\left(\boldsymbol{a}_o; \Sigma(\boldsymbol{a}_o)\left(\sum_{m=1}^{P}\mathbf{K}_m\widetilde{\boldsymbol{g}_o^{m\top}}\right), \left(\mathrm{diag}(\widetilde{\boldsymbol{\lambda}_o}) + \sum_{m=1}^{P}\mathbf{K}_m\mathbf{K}_m^\top\right)^{-1}\right). \tag{2}$$

The approximate posterior distribution of the projected instances can also be formulated as a product of multivariate normal distributions:

$$q(\{\mathbf{G}_o\}_{o=1}^{L}) = \prod_{o=1}^{L}\prod_{i=1}^{N}\mathcal{N}\left(\boldsymbol{g}_{o,i}; \Sigma(\boldsymbol{g}_{o,i})\left(\begin{bmatrix}\boldsymbol{k}_1^i\\ \vdots\\ \boldsymbol{k}_P^i\end{bmatrix}\widetilde{\boldsymbol{a}_o} + \widetilde{f_i^o}\widetilde{\boldsymbol{e}} - \widetilde{b_o}\widetilde{\boldsymbol{e}}\right), \left(\mathbf{I} + \widetilde{\boldsymbol{e}\boldsymbol{e}^\top}\right)^{-1}\right) \tag{3}$$

where the kernel weights and the auxiliary variables defined for each label are used together.

The approximate posterior distributions of the priors on the biases and the kernel weights can be found as products of gamma distributions:

$$q(\boldsymbol{\gamma}) = \prod_{o=1}^{L} \mathcal{G}\left(\gamma_o; \alpha_\gamma + \frac{1}{2}, \left(\frac{1}{\beta_\gamma} + \frac{\widetilde{b_o^2}}{2}\right)^{-1}\right) \tag{4}$$

$$q(\boldsymbol{\omega}) = \prod_{m=1}^{P} \mathcal{G}\left(\omega_m; \alpha_\omega + \frac{1}{2}, \left(\frac{1}{\beta_\omega} + \frac{\widetilde{e_m^2}}{2}\right)^{-1}\right). \tag{5}$$

The approximate posterior distribution of the biases and the kernel weights is a product of multivariate normal distributions:

$$q(\boldsymbol{b}, \boldsymbol{e}) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{b} \\ \boldsymbol{e}\end{bmatrix}; \Sigma(\boldsymbol{b}, \boldsymbol{e}) \begin{bmatrix} \mathbf{1}^\top \widetilde{\boldsymbol{f}^{1}}^\top \\ \vdots \\ \mathbf{1}^\top \widetilde{\boldsymbol{f}^{L}}^\top \\ \sum_{o=1}^{L} \widetilde{\mathbf{G}_o} \widetilde{\boldsymbol{f}^{o}}^\top \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} \widetilde{\gamma_1} + N & \dots & 0 & \mathbf{1}^\top \widetilde{\mathbf{G}_1^\top} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \widetilde{\gamma_L} + N & \mathbf{1}^\top \widetilde{\mathbf{G}_L^\top} \\ \widetilde{\mathbf{G}_1}\mathbf{1} & \dots & \widetilde{\mathbf{G}_L}\mathbf{1} & \mathrm{diag}(\widetilde{\boldsymbol{\omega}}) + \sum_{o=1}^{L} \widetilde{\mathbf{G}_o \mathbf{G}_o^\top} \end{bmatrix}^{-1} \right). \tag{6}$$

The approximate posterior distribution of the auxiliary variables is a product of truncated normal distributions:

$$q(\mathbf{F}) = \prod_{o=1}^{L} \prod_{i=1}^{N} \mathcal{TN}(f_i^o; \widetilde{\boldsymbol{e}^\top} \widetilde{\boldsymbol{g}_{o,i}} + \widetilde{b_o}, 1, f_i^o y_i^o > \nu) \tag{7}$$

where we need to find the posterior expectations in order to update the approximate posterior distributions of the projected instances and the classification parameters. Fortunately, the truncated normal distribution has a closed-form formula for its expectation.

3.2. Complete Algorithm

The complete inference algorithm is listed in Algorithm 1. The inference mechanism sequentially updates the approximate posterior distributions of the model parameters and the latent variables until convergence, which can be checked by monitoring the lower bound. The first term of the lower bound corresponds to the sum of exponential forms of the distributions in the joint likelihood. The second term is the sum of negative entropies of the approximate posteriors in the ensemble. The only nonstandard distribution in the second term is the truncated normal distributions of the auxiliary variables; nevertheless, the truncated normal distribution has a closed-form formula also for its entropy.

Algorithm 1 Bayesian Multilabel Multiple Kernel Learning

Require: $\{\mathbf{K}_m\}_{m=1}^{P}$, $\mathbf{Y}$, ν, α_γ, β_γ, α_λ, β_λ, α_ω, and β_ω

```
Initialize q(A), q({G_o}_{o=1}^L), q(b, e), and q(F) randomly
repeat
    Update q(Λ) and q(A) using (1) and (2)
    Update q({G_o}_{o=1}^L) using (3)
    Update q(γ), q(ω), and q(b, e) using (4), (5), and (6)
    Update q(F) using (7)
until convergence
return q(A) and q(b, e)
```

3.3. Prediction

In the prediction step, we can replace $p(\mathbf{A}|\{\mathbf{K}_m\}_{m=1}^{P}, \mathbf{Y})$ with its approximate posterior distribution $q(\mathbf{A})$ and obtain the predictive distribution of the intermediate outputs $\{\boldsymbol{g}_{o,\star}\}_{o=1}^{L}$ for a new data point as

$$p(\{\boldsymbol{g}_{o,\star}\}_{o=1}^{L}|\{\boldsymbol{k}_{m,\star}, \mathbf{K}_m\}_{m=1}^{P}, \mathbf{Y}) = \prod_{o=1}^{L}\prod_{m=1}^{P} \mathcal{N}(g_{o,\star}^{m}; \mu(\boldsymbol{a}_o)^\top \boldsymbol{k}_{m,\star}, 1 + \boldsymbol{k}_{m,\star}^\top \Sigma(\boldsymbol{a}_o)\boldsymbol{k}_{m,\star}).$$

The predictive distribution of the auxiliary variables $\boldsymbol{f}_\star$ can also be found by replacing $p(\boldsymbol{b}, \boldsymbol{e} | \{\mathbf{K}_m\}_{m=1}^P, \mathbf{Y})$ with its approximate posterior distribution $q(\boldsymbol{b}, \boldsymbol{e})$:

$$p(\boldsymbol{f}_\star | \{\boldsymbol{g}_{o,\star}\}_{o=1}^L, \{\mathbf{K}_m\}_{m=1}^P, \mathbf{Y}) = \prod_{o=1}^{L} \mathcal{N}\left(f_\star^o; \mu(b_o, \boldsymbol{e})^\top \begin{bmatrix} 1 \\ \boldsymbol{g}_{o,\star} \end{bmatrix}, 1 + \begin{bmatrix} 1 & \boldsymbol{g}_{o,\star} \end{bmatrix} \Sigma(b_o, \boldsymbol{e}) \begin{bmatrix} 1 \\ \boldsymbol{g}_{o,\star} \end{bmatrix} \right)$$

and the predictive distribution of the class label $\boldsymbol{y}_\star$ can be formulated using the auxiliary variable distribution:

$$p(y_\star^o = +1 | \{\boldsymbol{k}_{m,\star}, \mathbf{K}_m\}_{m=1}^P, \mathbf{Y}) = (\mathcal{Z}_\star^o)^{-1} \Phi\left(\frac{\mu(f_\star^o) - \nu}{\Sigma(f_\star^o)} \right) \quad \forall o$$

where $\mathcal{Z}_\star^o$ is the normalization coefficient calculated for the test data point and $\Phi(\cdot)$ is the standardized normal cumulative distribution function.

4. Experiments

In this section, we present the experimental results using our proposed Bayesian MKL algorithm in two different scenarios: affective image classification and affective image retrieval. We implemented our method in Matlab and took 200 variational iterations for inference with non-informative priors. We calculated the standard Gaussian kernel on each feature representation separately and picked the kernel width as $2\sqrt{D_m}$, where D_m is the dimensionality of corresponding feature representation.

4.1. Data Sets

Two affective image data sets have been used in the experiments, the `International Affective Picture System (IAPS)` [30] and the `ArtPhoto` [4].

The `IAPS` data set is a widely-used stimulus set in emotion-related studies. It contains altogether 1182 color images that cover contents across a large variety of semantic categories, including snakes, insects, animals, landscapes, babies, guns, and accidents, among others. Each image is evaluated by subjects (males & females) on three continuously varying scales from 1 to 9 for Valence, Arousal, and Dominance. A subset of 394 `IAPS` images have been grouped into 8 discrete emotional categories based on a psychophysical study [21]. Among the 8 emotions, Amusement, Awe, Contentment, and Excitement are considered as the positive class, whereas Anger, Disgust, Fear, and Sad are considered as the negative class. The ground truth label for each image was selected as the category that had majority of the votes. Both Machajdik *et al.* [4] and Lu *et al.* [10] used this subset for emotional image classification, and we also used it in our experiment to compare with their results.

The `ArtPhoto` data set was originally collected by Machajdik *et al.* [4] and it contains 806 artistic photographs obtained using discrete emotional categories as search queries in a photo sharing site. The discrete categories are the same as those adopted in the above `IAPS` subset and the images cover a wide range of semantic contents as well. We used this data set in our image retrieval experiment.

4.2. Image Features

We have used a set of ten low-level content descriptors for still images, including color, shape, and texture features. Four of them are standard MPEG-7 [31] visual descriptors: Scalable Color, Dominant Color, Color Layout, and Edge Histogram. These low-level features have been widely used in image classification and retrieval tasks, as well as in image affect detections [13, 14, 8]. When presented a new picture or painting, people tend to first get a holistic impression of it and then go into segments and details [32]. Therefore, our features are extracted both globally and locally from each image. For certain features, a five-zone image partitioning scheme (see Figure 4) is applied prior to the feature extraction [33]. Similar to the rule of thirds used in photography, the central part of an image usually catches most of people's attention. All the features have been extracted by using PicSOM system [34]. Table 2 gives a summary of these features.

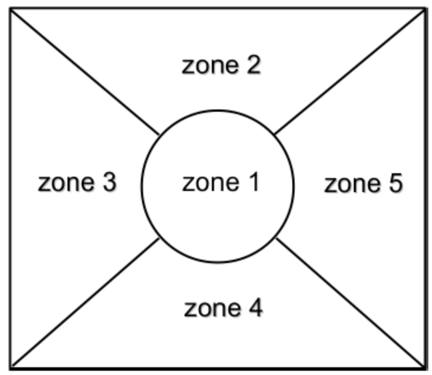


Figure 4: The five-zone partitioning scheme [33].

Table 2: The set of low-level image features used.

Index	Feature	Type	Zoning	Dims.
F1	Scalable Color	Color	Global	256
F2	Dominant Color	Color	Global	6
F3	Color Layout	Color	8×8	12
F4	5Zone-Color	Color	5	15
F5	5Zone-Colm	Color	5	45
F6	Edge Histogram	Shape	4×4	80
F7	Edge Fourier	Shape	Global	128
F8	5Zone-Edgehist	Shape	5	20
F9	5Zone-Edgecoocc	Shape	5	80
F10	5Zone-Texture	Texture	5	40

4.2.1. Color Features

Scalable Color: The descriptor is a 256-bin color histogram in HSV color space, which is encoded by a Haar transform.

Dominant Color: The descriptor is a subset from the original MPEG-7

XM descriptor and is composed of the LUV color system values of the first and second most dominant color. If the XM routine only found one dominant color, then it was duplicated.

Color Layout: The image area is divided in 8×8 non-overlapping blocks where the dominant colors are solved in YCbCr $(6, 3, 3)$ color system. Discrete Cosine Transform (DCT) is then applied to the dominant colors in each channel and the coefficients of DCT used as a descriptor.

5Zone-Color: This descriptor is a three-element vector that contains the average RGB values of all the pixels within each zone.

5Zone-Colm: The color moments feature treats the HSV color channels from each zone as probability distributions, and calculates the first three moments (mean, variance, and skewness) for each distribution.

4.2.2. Shape Features

Edge Histogram: The image is divided in 4×4 non-overlapping sub-images where the relative frequencies of five different edge types (vertical, horizontal, $45°$, $135°$, non-directional) are calculated using 2×2-sized edge detectors for the luminance of the pixels. The descriptor is obtained with a nonlinear discretization of the relative frequencies.

Edge Fourier: This descriptor calculates the magnitude of the 16×16 Fast Fourier Transform (FFT) of Sobel edge image.

5Zone-Edgehist: The edge histogram feature is the histogram of four Sobel edge directions. It is not the same as the MPEG-7 descriptor with the same name.

5Zone-Edgecoocc: The edge co-occurrence gives the co-occurrence matrix of four Sobel edge directions.

4.3. Texture Features

5Zone-Texture: The texture neighborhood feature is calculated from the Y (luminance) component of the YIQ color representation of each zone pixels. The 8-neighborhood of each inner pixel is examined, and a probability estimate is calculated for the probabilities that the neighbor pixel in each surrounding relative position is brighter than the central pixel. The feature vector contains these eight probability estimates.

4.4. Affective Image Classification

4.4.1. Experimental Setup

In this experiment, we evaluate the performance of the proposed Bayesian MKL algorithm within a classification framework and compare with the results in [4] and [10]. Note that for [10], we compared the result by using their proposed image shape features. The `IAPS` subset was used in this task. For training and testing, we used the same procedure as in [4, 10]: we used 5-fold Cross-Validation (CV) and calculated the average classification accuracy. As a baseline method, the standard SVM (with Gaussian kernel and 5-fold CV) was also implemented for comparison, where each feature was taken separately for training a single classifier. As for the free parameters, we manually set $\{\nu, \alpha_\gamma, \beta_\gamma, \alpha_\lambda, \beta_\lambda, \alpha_\omega, \beta_\omega\}$ to be $\{1, 1, 1, 0.001, 0.001, 0.001, 0.001\}$ respectively, based on the cross-validation results from the training data set. Through our experiments, we found that the last four parameters $\{\alpha_\lambda, \beta_\lambda, \alpha_\omega, \beta_\omega\}$ need careful selections as they directly control the kernels or features sparsity, whereas the other three ones do not affect the final performance much on emotional image predictions.

4.4.2. Results

Figure 5 shows the classification results (average of 8 classes). It is clear to see that our proposed algorithm is the best among the three. With rather generic low-level image features, our classifier can achieve better classification performance than methods of [4, 10] which rely on the design of complicated domain-specific features. Table 3 shows the comparison result with four other existing MKL methods, including the RBMKL [35], GMKL [36], NLMKL [37], and GLMKL [38]. The same ten low-level image features described in this article were utilized as the input for all the MKL methods. We can see that

Table 3: The comparison of classification accuracy between the proposed Bayesian MKL algorithm and four other existing MKL methods [35, 36, 37, 38].

Bayesian MKL	RBMKL	NLMKL	GMKL	GLMKL
0.31	0.24	0.29	0.26	0.29

our method is slightly better than NLMKL and GLMKL, yet much better than RBMKL and GMKL.

To further demonstrate the advantage of multiple kernel (multiview) learning over single kernel (single-view) learning, we trained and tested a single SVM classifier using each of the 10 features separately (with the same partition as MKL setup). Table 4 lists the classification accuracies. The best SVM classifier (trained with Dominant Color) can only achieve an accuracy of 22%, which is about 9 percent lower than that of our algorithm. And an SVM using all 10 features can give an accuracy of 25%. This demonstrates the advantage of multiview learning over single-view learning. It also validates the strength of our proposed classifier in terms of mapping low-level image features to high-level emotional responses.

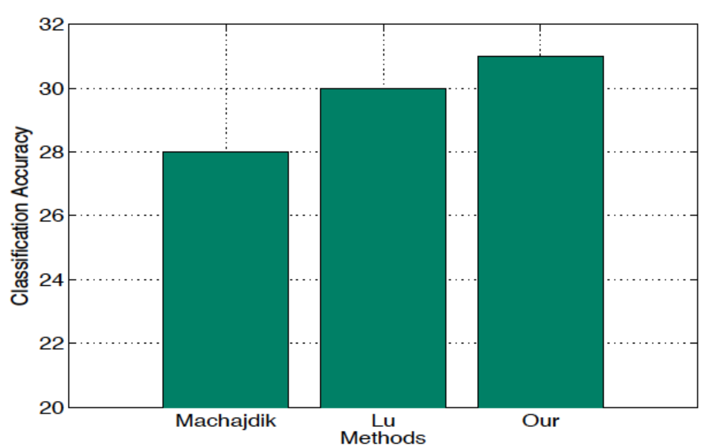


Figure 5: The classification results of the compared methods.

Table 4: The image features ranked by SVM classification accuracies.

Rank	Feature	Accuracy
1	Dominant Color	0.22
2	Color Layout	0.22
3	Edge Fourier	0.22
4	5Zone-Texture	0.21
5	5Zone-Colm	0.21
6	Scalable Color	0.20
7	5Zone-Color	0.20
8	5Zone-Edgecoocc	0.20
9	5Zone-Edgehist	0.19
10	Edge Histogram	0.18

We also compared the computational cost between the proposed Bayesian MKL algorithm and the single-feature method (used in [4, 10]), i.e., an algorithm based on the classification performance of a single feature at a time and selecting only those features which resulted in an average performance better than a pre-defined threshold. The classifier used in the single-feature method is SVM, which has been widely-utilized in emotional image classification tasks (e.g. [13, 15, 7, 16, 8]). Table 5 lists the compared computational costs measured in seconds. We can see that our method costs only around 1/4 of the time needed by the *single-feature + SVM* approach (with the best classification accuracy reaching 28%). Clearly, the single-feature approach is computationally much heavier than our method as it has to test each single feature first and then select the best features subset for the final (SVM) classifier input.

Another advantage of our MKL algorithm is that it can automatically assign weights to features without explicit feature extraction and selection procedures. Figure 6 shows the average feature representation weights (i.e., kernel weights) in the range $[0, 1]$ based on 5-fold CV for the multiple kernel learning scenario.

Table 5: The comparison of computational cost (in seconds) between the proposed Bayesian MKL and the single-feature + SVM method. We selected the SVM kernel width using a grid search within the set $\{0.01, 0.1, 1, 1.5, 2, 10, 100, 1000, 10000\}$ based on cross-validation results.

Bayesian MKL	Single-feature + SVM
557.80	2063.81

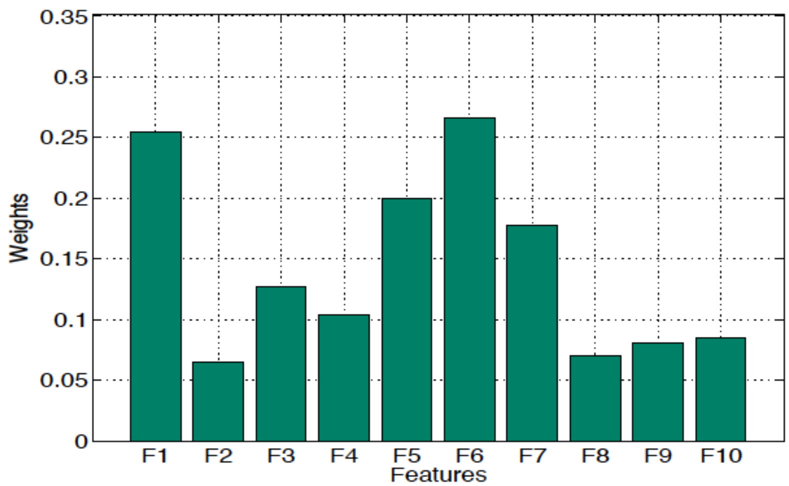
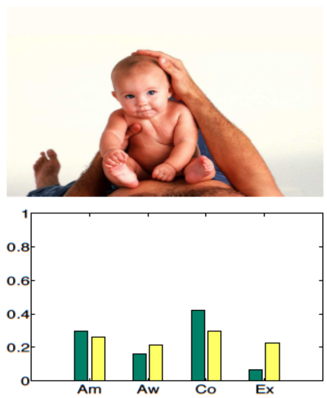


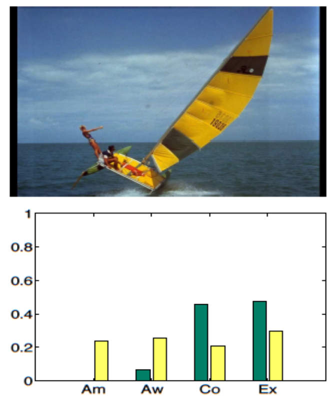
Figure 6: The average feature representation weights over 5-fold cross-validation for the multilabel multiple kernel learning scenario.

We clearly see that, among the ten image feature representations, Edge Histogram (F6) ranks first, followed by Scalable Color (F1), 5Zone-Colm (F5), and Edge Fourier (F7) etc. This reveals that colors and edges of an image are the most informative features for emotions recognition, which is in complete agreement with the studies in [4] and [10]. This also shows that multiple kernel learning helps to identify the relative importance of feature representations using a common set of kernel weights.

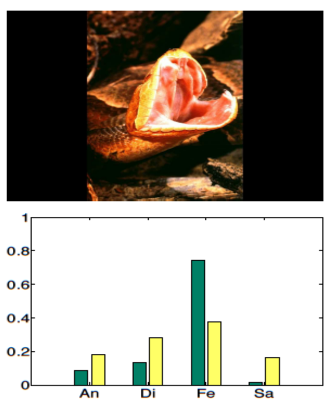
It is worth emphasizing that an image can evoke mixed emotions instead of a single emotion. Our Bayesian classifier is capable of producing multiple probabilistic outputs simultaneously, which allows us to give a “soft” class assignment



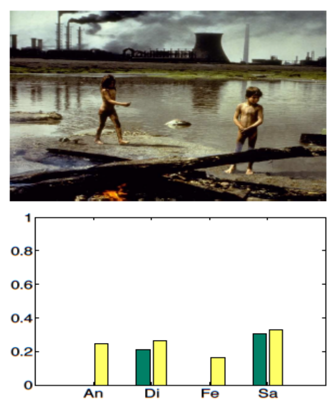
(a) Contentment



(b) Excitement



(c) Fear



(d) Sad

Figure 7: The agreement of image emotion distribution between our predicted results (green bars) and the normalized human votes (yellow bars). The x-axis shows positive emotions ((a) & (b)): Amusement, Awe, Contentment, Excitement, and negative emotions ((c) & (d)) Anger, Disgust, Fear, Sad. The y-axis shows the agreement in the range $[0, 1]$.

instead of a "hard" one. This characteristic is particularly useful for detecting emotion distribution evoked by an image. Figure 7 gives some examples. One can see that the probabilistic outputs of our Bayesian algorithm generally agree well with the real human votes for certain images.

4.5. Affective Image Retrieval

4.5.1. Experimental Setup

Also, we have designed an experiment for affective image retrieval based on our proposed Bayesian MKL algorithm. Firstly, we define the dissimilarity measure (the Euclidean distance in the implicit feature space) between a query image ($\boldsymbol{q}$) and a retrieved image ($\boldsymbol{r}$) as

$$d_e(\boldsymbol{q}, \boldsymbol{r}) = \sqrt{k_e(\boldsymbol{q}, \boldsymbol{q}) + k_e(\boldsymbol{r}, \boldsymbol{r}) - 2k_e(\boldsymbol{q}, \boldsymbol{r})}$$

$$k_e(\boldsymbol{q}, \boldsymbol{q}) = \sum_{m=1}^{P} e_m k_m(\boldsymbol{q}, \boldsymbol{q})$$

$$k_e(\boldsymbol{r}, \boldsymbol{r}) = \sum_{m=1}^{P} e_m k_m(\boldsymbol{r}, \boldsymbol{r})$$

$$k_e(\boldsymbol{q}, \boldsymbol{r}) = \sum_{m=1}^{P} e_m k_m(\boldsymbol{q}, \boldsymbol{r})$$

where $k_m(\cdot, \cdot)$ denotes the kernel function calculated on the mth feature representation and e_m is the weight for the corresponding kernel learned by our algorithm. Therefore, given a query image $\boldsymbol{q}$, our aim is to find those images with the smallest $d_e(\boldsymbol{q}, \boldsymbol{r})$ values. In essence, the smaller $d_e(\boldsymbol{q}, \boldsymbol{r})$ is, the more probable that the retrieved image $\boldsymbol{r}$ evokes similar emotional states in people. We selected query images from the `ArtPhoto` data set and let the algorithm retrieve images from the `IAPS` data set. Both data sets use the same emotional categories. The kernel weights $\{e_m\}_{m=1}^{P}$ were selected by training on the whole `IAPS` data set. Note that neither of the compared methods [4, 10] had explored image emotions from a retrieval perspective as their focus was on feature design only.

4.5.2. Results

Figure 8 gives some query-return examples from the results of image retrieval experiments. For the "Contentment" image, our algorithm successfully finds three other contentment images as its nearest neighbors. Similar query-return patterns can be seen from the "Disgust" and "Fear" query images. An interesting phenomenon is that both the 'Amusement" and "Excitement" query images have retrieved the “Awe” image, and both the “Anger” and “Sad” query images have found the “Fear” image among their top candidates. This is meaningful in that the former three emotions belong to the positive class which usually induces high valence, while the latter three emotions belong to the negative class which usually induces low valence but high arousal. Besides, the retrieval result again reveals the fact that an image often evokes multiple emotional states that are correlated with each other. For example, an amusement image usually elicits partial feeling of awe, and the feeling of sadness is closely connected with the feeling of fear. To a certain extent, our algorithm has detected such correlations that exist among emotions using rather low-level image features.

5. Conclusions

In this paper, we have presented a novel Bayesian multiple kernel learning algorithm for affective image classification and retrieval tasks with multiple outputs and feature representations. Instead of single feature (view) representation, our method adopts a kernel-based multiview learning approach for better prediction performance and interpretation, with the advantage of selecting or ranking features automatically. To capture the correlations between emotions, our method has been implemented within a multilabel setup. Due to its probabilistic nature, the proposed algorithm is able to predict a set of emotions evoked by an image rather than a single one. Currently, only the conventional low-level image features are utilized, as our focus in this paper is not on the affective feature design. Rather, we would like to provide a new framework for better predicting people’s emotional states, especially when an image evokes multiple affective feelings in people.

It is worth emphasizing that our method is not confined to the image emotions recognition, but can be easily extended to other affective stimuli such as audio and video data. Due to the varying subjectivity in humans and the limit of the available affective databases, it is of course not guaranteed that our method can make a perfect classification or retrieval for every single image. Eventually,

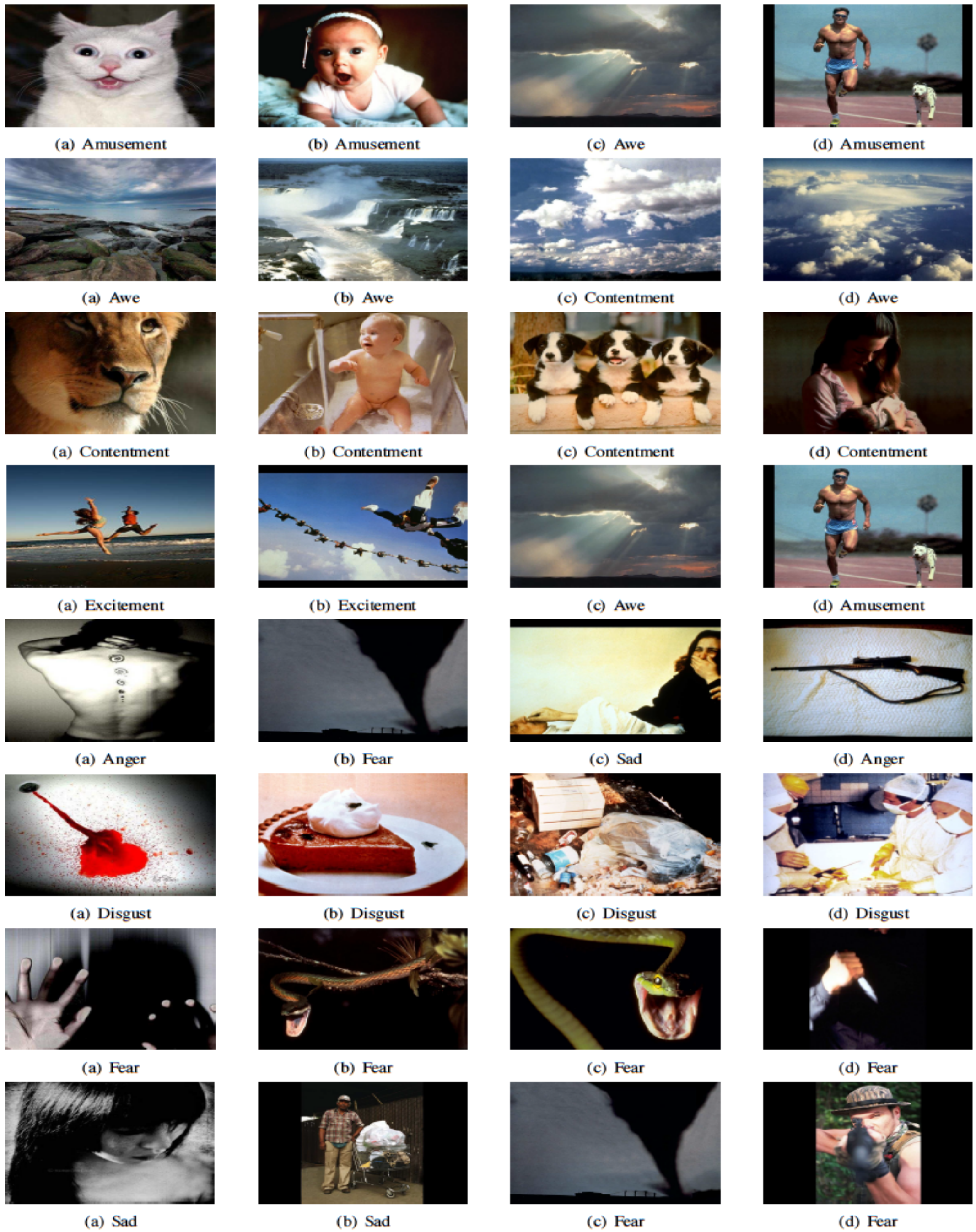


Figure 8: The image retrieval results using the `ArtPhoto` images as queries. The first column corresponds to the query images from the `ArtPhoto` data set, and the last three columns correspond to the top three retrieved images from the `IAPS` emotional subset ranked by distance. The ground-truth label is given under each image.

the development in this interdisciplinary area relies on the joint efforts from artificial intelligence, computer vision, pattern recognition, cognitive science, psychology, as well as color and art theories.

Acknowledgments

This work was financially supported by the Academy of Finland (Finnish Centre of Excellence in Computational Inference Research COIN, grant no 251170).

References

[1] R. Picard, Affective Computing, MIT Press, 1997.

[2] R. Datta, D. Joshi, J. Li, J. Wang, Image retrieval: Ideas, influences, and trends of the new age, ACM Computing Surveys 40 (2) (2008) 1–60.

[3] W. Wang, Q. He, A survey on emotional semantic image retrieval, in: Proceedings of International Conference on Image Processing, 2008, pp. 117–120.

[4] J. Machajdik, A. Hanbury, Affective image classification using features inspired by psychology and art theory, in: Proceedings of International Conference on Multimedia, 2010, pp. 83–92.

[5] L. Ou, M. Luo, A. Woodcock, A. Wright, A study of colour emotion and colour preference. Part I: Colour emotions for single colours, Color Research & Application 29 (3) (2004) 232–240.

[6] A. Hanjalic, Extracting moods from pictures and sounds: Towards truly personalized TV, IEEE Signal Processing Magazine 23 (2) (2006) 90–100.

[7] W. Wang, Y. Yu, S. Jiang, Image retrieval by emotional semantics: A study of emotional space and feature extraction, in: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2006, pp. 3534–3539.

[8] H. Zhang, E. Augilius, T. Honkela, J. Laaksonen, H. Gamper, H. Alene, Analyzing emotional semantics of abstract art using low-level image features, in: Proceedings of International Conference on Advances in Intelligent Data analysis, 2011, pp. 413–423.

[9] C. Colombo, A. Del Bimbo, P. Pala, Semantics in visual information retrieval, IEEE Multimedia 6 (3) (1999) 38–53.

[10] X. Lu, P. Suryanarayan, R. B. Adams Jr, J. Li, M. Newman, J. Wang, On shape and the computability of emotions, in: Proceedings of International Conference on Multimedia, ACM, 2012, pp. 229–238.

[11] C. Li, T. Chen, Aesthetic visual quality assessment of paintings, IEEE Journal of Selected Topics in Signal Processing 3 (2) (2009) 236–252.

[12] N. Bianchi-Berthouze, K-DIME: An affective image filtering system, IEEE Multimedia 10 (3) (2003) 103–106.

[13] Q. Wu, C. Zhou, C. Wang, Content-based affective image classification and retrieval using support vector machines, in: Affective Computing and Intelligent Interaction, Springer, 2005, pp. 239–247.

[14] L. Shamir, T. Macura, N. Orlov, D. Eckley, I. Goldberg, Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art, ACM Transactions on Applied Perception 7 (2010) 1–17.

[15] R. Datta, D. Joshi, J. Li, J. Wang, Studying aesthetics in photographic images using a computational approach, in: Proceedings of European Conference on Computer Vision, 2006, pp. 288–301.

[16] V. Yanulevskaya, J. Van Gemert, K. Roth, A. Herbold, N. Sebe, J. Geusebroek, Emotional valence categorization using holistic image features, in: Proceedings of International Conference on Image Processing, IEEE, 2008, pp. 101–104.

[17] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: Proceedings of Annual Conference on Computational Learning Theory, 1998, pp. 92–100.

[18] M. Gönen, E. Alpaydın, Multiple kernel learning algorithms, Journal of Machine Learning Research 12 (Jul) (2011) 2211–2268.

[19] C. Osgood, G. Suci, P. Tannenbaum, The Measurement of Meaning, University of Illinois Press, 1957.

[20] J. Russell, A. Mehrabian, Evidence for a three-factor theory of emotions, Journal of Research in Personality 11 (3) (1977) 273–294.

[21] J. Mikels, B. Fredrickson, G. Larkin, C. Lindberg, S. Maglio, P. Reuter-Lorenz, Emotional category data on images from the International Affective Picture System, Behavior Research Methods 37 (4) (2005) 626–630.

[22] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297.

[23] I. Rish, An empirical study of the naive Bayes classifier, in: Proceedings of IJCAI-01 Workshop on Empirical Methods in AI, 2001, pp. 41–46.

[24] L. Breiman, Classification and Regression Trees, Chapman & Hall/CRC, 1984.

[25] M. Boutell, J. Luo, X. Shen, C. Brown, Learning multi-label scene classification, Pattern Recognition 37 (9) (2004) 1757–1771.

[26] M. Gönen, Bayesian efficient multiple kernel learning, in: Proceedings of International Conference on Machine Learning, 2012, pp. 1–8.

[27] J. H. Albert, S. Chib, Bayesian analysis of binary and polychotomous response data, Journal of the American Statistical Association 88 (422) (1993) 669–679.

[28] N. Lawrence, M. Jordan, Semi-supervised learning via Gaussian processes, in: Advances in Neural Information Processing Systems, 2005, pp. 753–760.

[29] M. J. Beal, Variational algorithms for approximate Bayesian inference, Ph.D. thesis, The Gatsby Computational Neuroscience Unit, University College London (2003).

[30] P. Lang, M. Bradley, B. Cuthbert, International affective picture system (IAPS): Affective ratings of pictures and instruction manual, Tech. Rep. A-8, University of Florida, Gainesville, FL (2008).

[31] T. Sikora, The MPEG-7 visual standard for content description – An overview, IEEE Transactions on Circuits and Systems for Video Technology 11 (6) (2001) 696–702.

[32] R. Arnheim, Art and visual perception: A psychology of the creative eye, University of California Press, 1956.

[33] M. Sjöberg, H. Muurinen, J. Laaksonen, M. Koskela, PicSOM experiments in TRECVID 2006, in: Proceedings of the TRECVID 2006 Workshop, 2006, pp. 1–12.

[34] J. Laaksonen, M. Koskela, E. Oja, PicSOM-self-organizing image retrieval with MPEG-7 content descriptors, IEEE Transactions on Neural Networks 13 (4) (2002) 841–853.

[35] N. Cristianini, J. Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods, Cambridge university press, 2000.

[36] M. Varma, B. R. Babu, More generality in efficient multiple kernel learning, in: Proceedings of International Conference on Machine Learning, ACM, 2009, pp. 1065–1072.

[37] C. Cortes, M. Mohri, A. Rostamizadeh, Learning non-linear combinations of kernels., in: Advances in Neural Information Processing Systems, 2009, pp. 396–404.

[38] Z. Xu, R. Jin, H. Yang, I. King, M. R. Lyu, Simple and efficient multiple kernel learning by group lasso, in: Proceedings of International Conference on Machine Learning, 2010, pp. 1175–1182.