Identifying Emotions Aroused from Paintings

Xin Lu[1]([✉]), Neela Sawant[2], Michelle G. Newman[3],
Reginald B. Adams Jr.[3], James Z. Wang[3], and Jia Li[3]([✉])

[1] Adobe Systems Inc., Mountain View, USA
`xinl@adobe.com`
[2] Amazon.com, Inc., Seattle, USA
`neela.sawant@gmail.com`
[3] The Pennsylvania State University, State College, USA
`{mgn1,radams,jwang,jiali}@psu.edu`

Abstract. Understanding the emotional appeal of paintings is a significant research problem related to affective image classification. The problem is challenging in part due to the scarceness of manually-classified paintings. Our work proposes to apply statistical models trained over photographs to infer the emotional appeal of paintings. Directly applying the learned models on photographs to paintings cannot provide accurate classification results, because visual features extracted from paintings and natural photographs have different characteristics. This work presents an adaptive learning algorithm that leverages labeled photographs and unlabeled paintings to infer the visual appeal of paintings. In particular, we iteratively adapt the feature distribution in photographs to fit paintings and maximize the joint likelihood of labeled and unlabeled data. We evaluate our approach through two emotional classification tasks: distinguishing positive from negative emotions, and differentiating reactive emotions from non-reactive ones. Experimental results show the potential of our approach.

Keywords: Classification · Evoked emotion · Adaptive learning · Photograph · Visual art

1 Introduction

Visual artworks such as paintings can evoke a variety of emotional responses from human observers, such as calmness, dynamism, turmoil, and happiness. Automatic inference of the emotions aroused from a given painting is an important research question due to its potential application in large-scale image management and human perception understanding. For instance, the affective capability of paintings might be leveraged to determine which artwork might be used to decorate workplaces, hospitals, gymnasia, and schools. The problem is highly

---

challenging because many paintings are abstract in nature. The exact association between visual features and evoked emotions is often not obvious.

An applicable framework that has been used to quantify general emotion recognition problem from color photographs [3,8,13,14] is to learn a statistical model that connects handcrafted visual features extracted from the training images with their associated emotional labels. However, unlike emotion recognition in photographs which can leverage existing annotated datasets such as the International Affective Picture System (IAPS) [10], we do not have a validated dataset with sufficient manually-labeled paintings. Previous methods [7,11,12] conducted training on a small collection (around a hundred pieces) of labeled paintings, which is insufficient and not publicly accessible. As the features of images from the same emotional category form a perplexing distribution in the feature space, a large labeled training dataset is needed to provide good coverage of possible variations. Establishing a large collection of paintings associated with emotional labels is time-consuming in that the subjectivity of visual appeal judgment to paintings requires the validation of the emotional labels to a collection of images.

One intuitive alternative is to apply those model learned from labeled photographs onto paintings straightforwardly. However, due to the difference in feature distributions between paintings and color photographs, as we will illustrate in Sect. 3, the statistics captured by the model is quite different from those in paintings. Experimental results (Sect. 5) also confirm that the model trained on photographs is inaccurate in recognizing emotions in paintings.
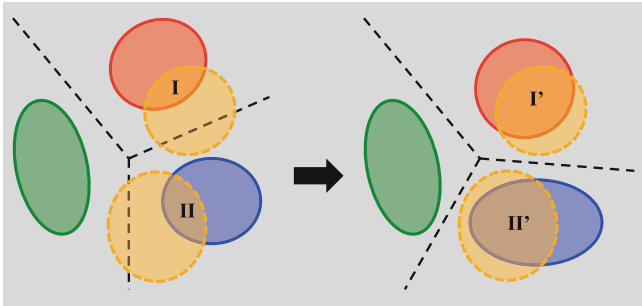

Fig. 1. Simplified illustration of distribution adaptation between photographs and paintings. **Left:** solid ellipses represent initially-estimated feature spaces of photographs from different emotional categories (indicated by different colors); orange dashed ellipses represent feature spaces of paintings whose emotional categories are unknown. The decision boundaries derived from photographs (black dashed lines) are unfit for paintings as they cut through feature spaces of paintings. **Right:** the estimation of photograph feature spaces adjusted according to the overlaps of photographs and paintings (region I and II). The new decision boundaries are more reasonable for paintings. (Color figure online)

This paper proposes an adaptive learning approach to recognize emotions in paintings, which leverages both labeled photographs and unlabeled paintings. The idea is to transfer the learned knowledge of photographs to paintings through distribution adaptation, a process wherein the distribution of the source domain is gradually adapted to the distribution of the target domain. Specifically, each photograph is associated with a weight; we account for the difference between the two distributions by re-weighting the weights. Figure 1 illustrates the basic intuition of this approach.

The rest of this paper is organized as follows: Sect. 2 provides a summary of related work. We present extensive statistical analysis to identify the dramatic distributions in paintings and color photographs in Sect. 3. The proposed algorithm is detailed in Sect. 4. Experimental results are presented in Sect. 5. Discussions and conclusions are provided in Sect. 6.

2   Related Work

2.1   Affective Image Classification

The analysis of emotions evoked through paintings has been under-explored by the research community, likely due to the scarcity of manually labeled paintings. Few studies have estimated aesthetics or emotions with a relatively small number of painting images [7,11,12]. Sartori *et al.* have studied abstract paintings using statistical analysis [16]. Our work is different in that we train statistical models on labeled photographs and adapt the learned models to paintings.

Some attempts were made to predict emotions from natural images [3,14,24] with psychologically validated labeled datasets (e.g., the IAPS). Commonly used visual features included color [2,22], texture [26], composition [25], and content of the image [15]. Machajdik and Hanbury [14] comprehensively modeled categorical emotions, using color, texture, composition, content, and semantic level features such as number of faces to model eight discrete emotional categories. Other representations of emotions that have also been explored by researchers include word pairs [18,23] and shape features [13]. As the relationship between these features and human emotions has been demonstrated on photographs, we believe these features also have indications to emotions aroused from paintings. In particular, our work adopted four groups of features: color, texture, composition, and content.

2.2   Domain Adaptation/Adaptive Learning

Many domain adaptation techniques have been developed in the past decades for building robust classifiers with data drawn from mismatched distributions. The two major directions are adapting feature distributions [6,17,20,21] and adapting classifier training [1,4,5].

To adapt feature distributions, Sugiyama *et al.* directly provided an estimate of the importance function by matching the two distributions in terms

of the Kullback-Leibler divergence [20]. Shi and Sha proposed an approach to learn domain-invariant features and use them to minimize a proxy misclassification error on the target domain [17]. Kang *et al.* [21] proposed an unsupervised domain adaptation approach where the classifier was trained iteratively, such that each iteration used an increased number of automatically discovered target domain examples, and a decreased number of source domain examples. Jhuo *et al.* [6] transformed the visual samples in the source domain into an intermediate representation such that each transformed source sample could be linearly reconstructed by the samples of the target domain. The intrinsic relatedness of the source samples was then captured by using a low-rank structure.

To build robust classifiers for data drawn from mismatched distributions, Bickel *et al.* [1] proposed a logistic regression classifier to explicitly model classification problems without having to estimate the marginal distributions for shift correction. Gopalan *et al.* [5] computed the domain shift by learning shared subspaces between the source and target domains for classifier training. In [9], joint bias and weight vectors were estimated as a max-margin optimization problem for domain adaptation. The authors of [4] enforced the target classifier to share similar decision values on the unlabeled consumer videos with the selected source classifiers.

Our work proposes an adaptive learning approach that integrates the feature adaptation and classifier training. We then leverages labeled photographs and unlabeled paintings to infer the visual appeal of paintings.

3   Feature Distributions in Paintings and Photographs

To better illustrate the problem and introduce the proposed adaptive learning algorithm, we first conduct statistical analyses to identify the differences of feature distributions between color photographs and paintings.

3.1   Settings

We analyzed the feature differences by taking the color photographs within the IAPS [10] and randomly crawling 10,000 paintings from Flickr.com. Photograph and painting examples are shown in Figs. 2 and 3.

We represent an image (photograph or painting) with five types of visual features: 21-dimensional global color features including statistics of saturation, brightness, hue, and colorfulness; 39-dimension region-based color features describing region-based color statistics; 27-dimensional texture features composed of wavelet textures and features that depict the contrast, correlation, and homogeneity for each of the HSV channels of the images; 14-dimensional feature encoding the depth of field, dynamics, and the rule of thirds to represent the composition of an image; and 4-dimensional content feature referring to the number and size of frontal faces and the number of skin-colored pixels. All dimensions of the feature vectors are normalized to $[0, 1]$. Detailed descriptions of those features are presented in [14].
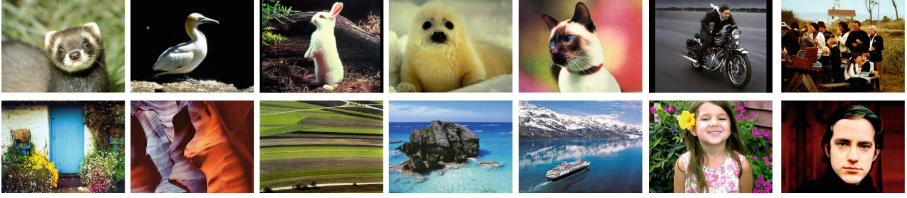
Fig. 2. Examples of photographs in the IAPS dataset [10]. (Color figure online)

Fig. 3. Examples of the painting images that we have collected for this study. (Color figure online)

3.2   Differences of Feature Distributions

This section unveils the underlying difference of feature distributions of paintings and photographs. We calculate the differences for each type of features using Euclidean distance as follows.

For each painting t from the set of paintings $T = \{t_i\}_{i=1}^{N_t}$ and its feature vector $f_c(t)$ ($c \in \{$color(global), color(region), texture, content, composition$\}$), we pair it with its nearest neighbor $S^*(t)$ from the photograph set $S = \{s_i\}_{i=1}^{N_s}$, where $S^*(t) = \arg\min_s D(f_c(t), f_c(s))$. N_s and N_t are the sizes of the photograph set and the painting set respectively. Distance $D(f_c(t), f_c(S^*(t)))$, denoted by $D_c(t)$, is defined as the distance between a single painting t and the collection of photographs $\{s_i\}$ in terms of feature type c. We normalize $D_c(t)$ by

$$\tilde{D}_c(t) = \frac{D_c(t)}{D(f_c(s'), f_c(S^*(t)))} \,, \tag{1}$$

where s' is the photo whose feature vector $f_c(s')$ is the nearest one to $f_c(S^*(t))$. $\tilde{D}_c(t) < 1$ means that the visual feature extracted from painting t is close to at least one feature vector in the photograph collection S, while $\tilde{D}_c(t) \geq 1$ indicates the existence of a larger difference between t's feature and one of the features from S. The greater $\tilde{D}_c(t)$ is, the larger the difference is between $f_c(t)$ and the photograph set S.

In Fig. 4, we show the distributions of the normalized distance $\tilde{D}_c$ between a feature vector (global color features, region-based color features, texture, composition, and content) in a painting and its nearest vector from the photograph set. As shown in the fourth plot, paintings differ from photographs most in terms
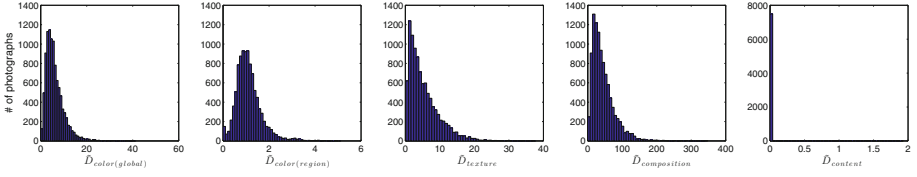
Fig. 4. Distributions of the normalized distance ($\tilde{D}$) from a painting to its nearest photograph, in terms of color (global), color (region), texture, composition and content feature, respectively. (Color figure online)
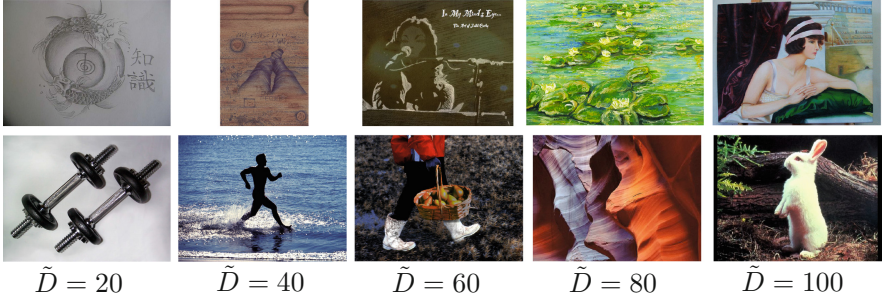

| $\tilde{D} = 20$ | $\tilde{D} = 40$ | $\tilde{D} = 60$ | $\tilde{D} = 80$ | $\tilde{D} = 100$ |

Fig. 5. Examples of painting-photograph pairs with different value of $\tilde{D}_{composition}$. The first row are paintings. Their associated photographs are in the second row.

of the composition; the value of $\tilde{D}_{composition}$ at the peak of the distribution is about 17. This indicates that there is dramatic differences in composition features between most paintings and photographs. Paintings and photographs also differ a lot in terms of the global color feature (first plot) and the texture feature (third plot), as their curves peak at $\tilde{D}_{color(global)}$ around 4 and $\tilde{D}_{texture}$ around 2, respectively. Finally, in the last plot, $\tilde{D}_{content}$ are close to 0 for almost all paintings, which indicates that photographs and paintings have similar content features. The reason may be that the content features we extracted only describe the existence and the number of human faces, as well as the size of human skin areas. The dramatic differences between feature distributions and paintings indicate the necessity to perform the proposed adaptive learning in order to leverage the labeled photographs for recognizing emotions in paintings.

In Figs. 5, 6 and 7, we provide some examples of painting-photograph pairs with different distances. Pairs with small $\tilde{D}_c$ are similar in terms of feature c.

4   Adaptive Learning Approach

We now introduce the detailed formulation of the proposed adaptive learning approach. We first explain the notations and provide a formal description of the common covariant shift approach mentioned in Sect. 2. We then present our

$$\tilde{D} = 2 \qquad \tilde{D} = 4 \qquad \tilde{D} = 6 \qquad \tilde{D} = 8 \qquad \tilde{D} = 10$$

Fig. 6. Examples of painting-photograph pairs with different value of $\tilde{D}_{texture}$. The first row are paintings. Their associated photographs are in the second row.

$$\tilde{D} = 3 \qquad \tilde{D} = 6 \qquad \tilde{D} = 9 \qquad \tilde{D} = 12 \qquad \tilde{D} = 15$$

Fig. 7. Examples of painting-photograph pairs with different value of $\tilde{D}_{color}$. The first row are paintings. Their associated photographs are in the second row.

approach that integrates the adaptive feature adaptation and classifier training. Finally, we describe how we jointly solve the maximization problem.

4.1   Notation

Let x be the p-dimensional data and the class labels of x be $y \in \{1, 2, \ldots, K\}$. For binary classification, K is set to two. Let S and T be the sets of photographs (source domain) and paintings (target domain), respectively, and the marginal probabilities $P_{X \in S}(X)$ and $P_{X \in T}(X)$ are denoted by $\Psi(x)$ and $\Phi(x)$, respectively. Let $\hat{\Phi}(x)$ and $\hat{\Psi}(x)$ denote the estimated distributions using the observed data samples.

4.2   Covariant Shift

Given the same feature observation $X = x$, the photograph set S and the painting set T, the conditional distributions of emotion labels Y are expected to be the same in both datasets, i.e., $P_{x \in S}(Y|X = x) = P_{x \in T}(Y|X = x)$. However, the marginal distributions of X may be different, i.e. $\Psi(X) \neq \Phi(X)$. This difference between the two domains is called covariate shift [19]. This is a problem

if a mis-specified statistical model from a parametric model family is trained by minimizing the expected classification error over S. A common covariate shift correction approach assigns fixed weights to each labeled instance in S proportional to the ratio $\frac{\Psi(X)}{\Phi(X)}$. Then a classifier $P(Y|X)$ is trained to minimize the weighted classification error. We call it static covariate shift correction, as the estimation of instance weights is fixed before the subsequent classifier training task.

4.3   Adaptive Learning Approach

We devise a semi-supervised adaptive learning algorithm using both labeled and unlabeled data. As in standard covariate shift correction approaches, we compute a weight $w(x) = \frac{\hat{\Phi}(x)}{\hat{\Psi}(x)}$ for each $x \in S$. Essentially $w(x)$ is a form of importance sampling where data from the photographs is selected with a weight that corrects the covariate shift in both photographs and paintings. Then, all labeled and unlabeled data can be treated in a common semi-supervised framework to maximize the following objective:

$$O = \sum_{(x,y) \in S \times Y} w(x)(\log P(x,y)) + \alpha \sum_{x' \in T} \log P(x') \, , \qquad (2)$$

where α is a pre-determined scaling factor associated with incomplete (unlabeled) data. In Eq. 2, $P(x') = \hat{\Phi}(x')$ and $P(x,y) = \hat{\Phi}(x)P(y|x)$. In the static way, $w(x)$ is estimated once as $\frac{\hat{\Phi}(x)}{\hat{\Psi}(x)}$ and then maintained constant throughout the optimization of Eq. 2. Such strategy does not incorporate any information from the consequent classification task. On the contrary, we update the weights in each iteration.

4.4   Mixture Discriminant Analysis

The iterative estimation of $P(x,y), x \in T$ and $\Phi(x)$ can be readily embodied in a semi-supervised framework using a mixture discriminant analysis (MDA). A K-class Gaussian mixture discriminant is computed as $P(X = x, Y = k) = a_k \sum_{r=1}^{R_k} \pi_{kr} \phi(x|\mu_{kr}, \Sigma_{kr})$, where a_k is the prior probability of class $k(0 \le a_k \le 1), \sum_{k=1}^{K} a_k = 1$. R_k is the number of mixture components used to model class k and the total number of mixture components for all the classes is $M = \sum_{k=1}^{K} R_k$. π_{kr} is the mixing proportion for the rth component in class k, $0 \le \pi_{kr} \le 1$, and $\sum_{k=1}^{K} \pi_{kr} = 1$. $\phi(.)$ denotes the pdf of a Gaussian distribution with μ_{kr} the centroids of component r in class k and σ_{kr} as the corresponding covariance matrix. To simplify the notation, the mixture model can be written as

$$P(X = x, Y = k) = \sum_{m=1}^{M} \pi_m p_m(k)\phi(x|\mu_m, \sigma_m) \, , \qquad (3)$$

where $1 \le m \le M$ is the new component label assigned in a consecutive manner to all the components in the classes. The prior probability for the mth component $\pi_m = a_k \pi_{kr}$ if m is the new label for the rth component in the kth class. The quantity $p_m(k) = 1$ if the component m belongs to class k and 0 otherwise. This ensures that the density of X within class k is a weighted sum over only the components inside class k.

Formulation of Joint Optimization. With weights initialized, we optimize Eq. 2 using expectation maximization algorithm with an intermediate classification step for the unlabeled examples in the paintings. Iterations are denoted by τ.

– *E-step*: Compute the posterior probability of each sample $(x, y) \in S \times Y$ belonging to component m.

$$q_m(x) \propto \pi_m^{(\tau)} p_m(y) \phi(x|\mu_m^{(\tau)}, \sigma_m^{(\tau)}), \quad \text{subject to} \sum_{m=1}^{M} q_m(x) = 1 . \quad (4)$$

For the unlabeled data $x' \in T$, the labels y' are to be treated as missing parameters. We first compute the posterior probability over each component m.

$$f_m(x') \propto \pi_m^{(\tau)} \phi(x|\mu_m^{(\tau)}, \sigma_m^{(\tau)}) . \quad (5)$$

Next, classification is conducted to estimate $y'^{(\tau)} = \arg\max_k \sum_{m \in \mathbb{R}_k} f_m(x')$. The quantity $p_m(y'^{(\tau)}) = 1$ and all other $p_{m' \ne m}(y'^{(\tau)}) = 0$. The posterior for unlabeled data is updated as:

$$q_m(x') \propto \pi_m^{(\tau)} p_m(y'^{(\tau)}) \phi(x'|\mu_m^{(\tau)}, \sigma_m^{(\tau)}), \quad \text{subject to} \sum_{m=1}^{M} q_m x' = 1 . \quad (6)$$

– *Maximization*: In this step, the parameters for paintings are updated using all data.

$$\pi_m^{(\tau+1)} \propto \sum_{x \in S} w^{(\tau)}(x) q_m(x) + \alpha \sum_{x' \in T} q_m(x'), \quad \text{subject to} \sum_m \pi_m^{(\tau+1)} = 1 . \quad (7)$$

$$\mu_{m,p}^{(\tau+1)} = \frac{\sum_{x \in S} w^{(\tau)}(x) q_m(x) x_p + \alpha \sum_{x' \in T} q_m(x') x'_p}{\sum_{x \in S} w^{(\tau)}(x) q_m(x) + \alpha \sum_{x' \in T} q_m(x')} . \quad (8)$$

Let

$$A = \sum_{x \in S} w^{(\tau)}(x) q_m(x) (x_p - \mu_{m,p}^{(\tau+1)})^2 , \quad (9)$$

$$B = \alpha \sum_{x' \in T} q_m(x')(x'_p - \mu_{m,p}^{(\tau+1)})^2 \, , \tag{10}$$

$$C = \sum_{x \in S} w^{(\tau)}(x)q_m(x) \, , D = \alpha \sum_{x' \in T} q_m(x') \, . \tag{11}$$

Then

$$\sigma_{m,p}^{2(\tau+1)} = \frac{A + B}{C + D} \, . \tag{12}$$

– *Weight Update*: Compute $P(X = x|Y = y), \forall(x,y) \in S$, using the updated parameters of class Y and update their weights as follows:

$$w(x)^{(\tau+1)} = \frac{\sum_m \pi_m p_m(y)\phi(x|y; \mu_m^{(\tau+1)}, \sigma_m^{2(\tau+1)})}{\hat{\Psi}(x)} \, . \tag{13}$$

In the above formulation, the parameters for unlabeled paintings (i.e., $\hat{\Psi}(x)$) always remain constant. Thus the adaptation is sensitive to the classification for paintings (the numerator), and weights are refined iteratively to consider both classification and clustering error.

5 Experiments

5.1 Settings

Datasets: We use three datasets: photograph dataset with emotional labels, unlabeled painting dataset, and a collection of 200 labeled paintings.

– *Labeled photographs:* We used the IAPS [10] as labeled photographs (Fig. 2). The IAPS dataset is a popular and validated dataset for the study of emotions evoked by natural photographs. The IAPS dataset contains $1,149$ images, each of which is associated with an empirically derived mean of valance and arousal. Valence describes the positive or negative aspect of human emotions, where common emotions, such as joy and happiness, are positive, whereas anger and fear are negative. Arousal represents the human physiological state of being
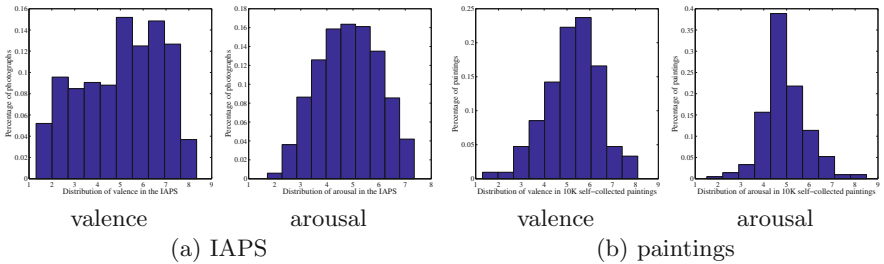

| valence | arousal | valence | arousal |
| (a) IAPS | | (b) paintings | |

Fig. 8. Distributions of valence and arousal in the IAPS dataset and the 10,000-painting dataset.

reactive to stimuli. A higher value of arousal indicates higher excitation. We generate the ground truth emotional ratings of the four classification tasks based on the value of valence and arousal of photographs. The range of valence in the IAPS is $[1.3, 8.3]$, and the range of arousal is $[1.7, 7.4]$. The distribution of valence and arousal in the IAPS is presented in Fig. 8(a).

– *Unlabeled paintings:* We randomly crawled $10,000$ paintings from Flickr as the unlabeled painting set. Examples have been presented in Sect. 3. A subset or a whole set of these paintings were used in our approach.
– *Labeled paintings:* We randomly crawled an alternative collection of paintings (200) from Flickr for the purpose of evaluation. We recruited participants to rate those paintings in terms of valence and arousal. The participants included college students with major in psychology and community individuals recruited from Amazon Mechanical Turk. Each painting was rated by at least five participants, and ratings were collected with the same guidelines as in the IAPS. The range of valence in rated paintings was $[1.3, 8.1]$, and the range of arousal was $[1.5, 8.5]$. The distribution of valence and arousal of labeled paintings is presented in Fig. 8(b).

Model selection and parameter tuning: To make it more convenient to introduce the tasks, we first briefly discuss the settings for the model selection and initialization.

– *Model selection:* We randomly selected 100 images from the labeled painting set as a validation set and used the remaining 100 paintings for test. We used a grid search to tune α and the number of unlabeled images to be used for semi-supervised learning using a validation dataset. Within each task, the number of mixture components (clusters) was determined using Bayesian Information Criterion (BIC). Several random initializations were evaluated to select a good model using the validation dataset.
– *Weight initialization:* We first approximated $\hat{\Phi}(x)$ and $\hat{\Psi}(x)$ by independently estimating Gaussian mixture models (ϕ) for the photograph domain and the painting domain. The initial weights of photograph domain data were computed by taking the ratio of $\hat{\Phi}(x)/\hat{\Psi}(x)$.

In the following three subsections, we present the settings and experimental results of the two classification tasks.

5.2   Classification Tasks and Results

We evaluated our approach with two emotion classification tasks. We first identified the positivity or negativity of emotion aroused from paintings. Then we analyze whether the emotional content in paintings was reactive or not. In both tasks, we compared the performance of our approach with the baseline approach in which the model was trained on labeled photographs and tested on paintings.

Task 1 - Identifying positivity and negativity of emotional content: As valence describes the positive or negative aspect of human emotions, we divided

paintings into two groups based on valence value. We calculated the mean value of valence in the IAPS, which was 5. Images with valence larger than 5 were labeled as positive (Class 1), and others were labeled as negative (Class 0). This results in 631 positive images and 514 negative images. In the validation set, there were 64 positive paintings and 36 negative ones. In the test set, 62 images were positive, and 38 were negative.

Task 2 - Identifying reactivity of emotional content: According to the psychology literature, the dimension of arousal refers to the human physiological state of being reactive to stimuli. We let images with arousal values larger than 4.8 as images with stronger reactive emotional content (Class 1) and lower than 4.8 has weaker reactive emotional content (Class 0). This results in 597 positive images and 551 negative images in training. In the validation set, there were 41 positive paintings and 59 negative ones. In the test set, 61 images were positive, and 39 were negative.

For both tasks, we compared our results with the baseline approach (MDA) in which the model was trained on labeled photographs and tested on paintings. Our approach outperformed the MDA approach in both the validation dataset and the test dataset for both tasks. For Task 1, the classification accuracy by MDA for the test dataset is 59 % (61 % for the validation dataset), while that

Fig. 9. Correctly classified and misclassified test paintings in the test set of the task 1. Paintings were annotated with TP, TN, FP, and FN, referring to correctly classified strongly reactive paintings, correctly classified weakly reactive paintings, misclassified strongly reactive paintings, and misclassified weakly reactive paintings, respectively. (Color figure online)

by our approach is 61 % (63 % for validation). For Task 2, the accuracy by MDA for the test dataset is 54 % (52 % for the validation dataset), while that by our approach is 61 % (62 % for validation).

Fig. 10. Correctly classified and misclassified paintings in the test set of the task 2. Paintings were annotated with TP, TN, FP, and FN, referring to correctly classified positive paintings, correctly classified negative paintings, misclassified positive paintings, and misclassified negative paintings, respectively. (Color figure online)

We show classification results on example images for the two tasks in Figs. 9 and 10. Abstract paintings with a strong visual difference from natural photographs tend to be misclassified by the learned model. This indicates that emotional responses evoked by similar stimuli (such as color and texture) might be different in natural photographs and abstract paintings. To better predict emotions aroused from abstract paintings, it is necessary to include labeled abstract paintings in the training set in addition to natural photographs. We also observe that some stimuli have different emotional indications in photographs and paintings. For instance, the color of blue is associated with negative emotions aroused from natural photographs, whereas the color of red and yellow are associated with positive emotions. However, this is not necessarily true in paintings as shown in Fig. 10. To improve the prediction accuracy on paintings in the wild, we may need to generalize the proposed algorithm in cases that we have some labeled paintings besides a large collection of labeled photographs and unlabeled paintings. We would like to take this direction as future work.

6   Discussions and Conclusions

We investigated the problem of emotion classification on paintings. Due to the scarcity of paintings with emotional labels, we proposed an adaptive learning approach that leveraged color photographs with emotion labels and unlabeled paintings to infer the emotional appeal of paintings. Our approach takes into account differences in feature distributions in paintings and color photographs as we use photographs with emotional ratings. We performed two emotion classification tasks. The experimental results showed that our approach achieved a higher accuracy in recognizing emotions in paintings.

Although we have shown that the adaptive learning approach improves clearly upon a baseline approach without adaption, the classification accuracies we achieved for classification of emotional responses are nevertheless low, indicating ample room for enhancement. We believe that the main reason for the limited performance is the intrinsic complexity of the problem. The visual features we have experimented with seem to have weak association with the evoked emotions of paintings, and it is quite possible that a fundamental breakthrough is needed to push further the technology. In addition, our adaptive learning approach relies on the assumption that the non-zero density support of the feature distribution of the source is the same as that of the target, under which reweighting is viable to approximate the distribution of the target. The validity of this assumption calls for thorough examination in the future.

References

1. Bickel, S., Brückner, M., Scheffer, T.: Discriminative learning for differing training and test distributions. In: International Conference on Machine Learning (ICML), pp. 81–88 (2007)
2. Changizi, M.A., Zhang, Q., Shimojo, S.: Bare skin, blood and the evolution of primate colour vision. Biol. Lett. **2**(2), 217–221 (2006)
3. Datta, R., Li, J., Wang, J.Z.: Algorithmic inferencing of aesthetics and emotion in natural image: an exposition. In: International Conference on Image Processing (ICIP), pp. 105–108 (2008)
4. Duan, L., Xu, D., Chang, S.F.: Exploiting web images for event recognition in consumer videos: a multiple source domain adaptation approach. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1338–1345 (2012)
5. Gopalan, R., Ruonan, L., Chellappa, R.: Domain adaptation for object recognition: an unsupervised approach. In: International Conference on Computer Vision (ICCV), pp. 999–1006 (2011)
6. Jhuo, I.H., Liu, D., Lee, D., Chang, S.F.: Robust visual domain adaptation with low-rank reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2168–2175 (2012)
7. Jia, J., Wu, S., Wang, X., Hu, P., Cai, L., Tang, J.: Can we understand van Gogh's mood? Learning to infer affects from images in social networks. In: ACM International Conference on Multimedia, pp. 857–860 (2012)
8. Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.T., Wang, J.Z., Li, J., Luo, J.: Aesthetics and emotions in images. IEEE Sig. Process. Mag. **28**(5), 94–115 (2011)

9. Khosla, A., Zhou, T., Malisiewicz, T., Efros, A.A., Torralba, A.: Undoing the damage of dataset bias. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7572, pp. 158–171. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33718-5_12

10. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system: affective ratings of pictures and instruction manual. In: Technical report A-8, University of Florida, Gainesville, FL (2008)

11. Li, C., Chen, T.: Aesthetic visual quality assessment of paintings. IEEE J. Sel. Top. Sig. Process. **3**(2), 236–252 (2009)

12. Li, C.T., Shan, M.K.: Emotion-based impressionism slideshow with automatic music accompaniment. In: ACM International Conference on Multimedia, pp. 839–842 (2007)

13. Lu, X., Suryanarayan, P., Adams Jr., R.B., Li, J., Newman, M.G., Wang, J.Z.: On shape and the computability of emotions. In: ACM International Conference on Multimedia, pp. 229–238 (2012)

14. Machajdik, J., Hanbury, A.: Affective image classification using features inspired by psychology and art theory. In: ACM International Conference on Multimedia, pp. 83–92 (2010)

15. Rodgers, S., Kenix, L.J., Thorson, E.: Stereotypical portrayals of emotionality in news photos. Mass Commun. Soc. **10**(1), 119–138 (2007)

16. Sartori, A., Yanulevskaya, V., Salah, A.A., Uijlings, J., Bruni, E., Sebe, N.: Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory. ACM Trans. Interact. Intell. Syst. **5**(2), 8 (2015)

17. Shi, Y., Sha, F.: Information-theoretical learning of discriminative clusters for unsupervised domain adaptation. In: International Conference on Machine Learning (ICML), pp. 1079–1086 (2012)

18. Shibata, T., Kato, T.: Kansei image retrieval system for street landscape-discrimination and graphical parameters based on correlation of two image systems. In: International Conference on Systems, Man, and Cybernetics, pp. 274–252 (2006)

19. Shimodaira, H.: Improving predictive inference under covariate shift by weighting the log-likelihood function. J. Stat. Plan. Infer. **90**(2), 227–244 (2000)

20. Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P.V., Kawanabe, M.: Direct importance estimation with model selection and its application to covariate shift adaptation. In: Neural Information Processing Systems (NIPS), pp. 1433–1440 (2008)

21. Tang, K., Ramanathan, V., Li, F.F., Koller, D.: Shifting weights: adapting object detectors from image to video. In: Neural Information Processing Systems (NIPS), pp. 647–655 (2012)

22. Valdez, P., Mehrabian, A.: Effects of color on emotions. J. Exp. Psychol. Gen. **123**(4), 394–409 (1994)

23. Wang, H.L., Cheong, L.F.: Affective understanding in film. IEEE Trans. Circ. Syst. Video Technol. **16**(6), 689–704 (2006)

24. Yanulevskaya, V., van Gemert, J., Roth, K., Herbold, A., Sebe, N., Geusebroek, J.: Emotional valence categorization using holistic image features. In: International Conference on Image Processing (ICIP), pp. 101–104 (2008)

25. Yao, L., Suryanarayan, P., Qiao, M., Wang, J.Z., Li, J.: Oscar: on-site composition and aesthetics feedback through exemplars for photographers. Int. J. Comput. Vis. **96**(3), 353–383 (2012)
26. Zhang, H., Augilius, E., Honkela, T., Laaksonen, J., Gamper, H., Alene, H.: Analyzing emotional semantics of abstract art using low-level image features. In: Advances in Intelligent Data Analysis, pp. 413–423 (2011)