Neural Processing Letters (2020) 51:2077–2086
https://doi.org/10.1007/s11063-019-10035-7

# Deep Transfer Learning for Image Emotion Analysis: Reducing Marginal and Joint Distribution Discrepancies Together

**Yuwei He**[1] · **Guiguang Ding**[1]

Published online: 22 April 2019


**Abstract**
A lot of research attentions have been paid to image emotion analysis in recent years. Meanwhile, as convolutional neural networks (CNNs) have made great successful in computer vision, many researchers start to employ CNN to discriminate image emotions. However, the training procedure of CNNs depends on sufficient labeled data. Therefore, a CNN is hard to perform well in an image domain with scant labeled information. In this paper, we propose a deep transfer learning method for image emotion analysis. The method can leverage rich emotion knowledge from a source domain to the target domain. Our method reduces both marginal and joint domain distribution discrepancies at fully-connected layers. Through this way, we can effectively extract more transferable features and advance the performance of CNNs on poor-label emotion-image domains.

**Keywords** Image emotion analysis · Transfer learning · Deep learning · Convolutional neural network

## 1 Introduction

Different visual content can evoke different human emotions, which directly influence our cognition and decision. Therefore, more researchers start to investigate and interpret human emotion contained in image content [30]. Most conventional methods design manually crafted features based on art and psychology theory and then recognize human emotions by discriminating these features [8,17,19,37].

Deep learning has made significant development in recent years, and the performance of convolutional neural networks (CNNs) on many computer vision tasks is comparable to that of humans. Meanwhile, large-scale image datasets boost feature learning based on CNNs. For example, a CNN pre-trained with ImageNet can extract more representative features for

✉ Yuwei He
hyw16@mails.tsinghua.edu.cn

Guiguang Ding
dinggg@tsinghua.edu.cn

[1] Tsinghua University School of Life Sciences, Beijing, People's Republic of China

general visual tasks. In the image emotion analysis field, studies have proved that CNN-based features are more discriminative compared with traditional manually crafted features [29].

However, there are still limitations for CNNs in image emotion analysis. Firstly, training CNN depends on massive labeled data. But in many emotion-image domains, the amount of labeled images is limited and manually labeling them is prohibitive [18]. Moreover, the scalability of CNNs is still limited as different emotion-image domains exhibit different image styles, which leads to different domain distributions. Therefore, even if a CNN performs well in an emotion-image domain, it may not achieve comparable performance in another one.

Transfer learning aims at transferring information from a rich-label source domain to another poor-label target domain [26]. The key technical problem is how to reduce the distribution discrepancy of the two domains. Recently deep transfer learning methods have been widely applied in computer visions [14,15,24,25]. One important reason is that a deep model can learn more domain-invariant features [29]. As deep models prefer to learn domain-specific features on top layers, the main bottleneck of deep transfer learning methods is to reduce the shift between two domain distributions of these layers.

In order to generalize CNNs to different emotion-image domains, in this paper, we design a novel deep transfer learning method to promote CNN-based emotion classifiers on small-scale image domains. Its advantages on image emotion analysis are as follows: 1) Our method requires two domains share the same CNN. As different emotion-image domains contain similarity elements on pixel-level, sharing the same CNN can learn higher quality image features at first-layers. 2) We have both considered marginal distribution discrepancy at the same layers [14] and joint distribution discrepancy of different layers [16]. The layers in deep models are trained jointly, so we should not only consider marginal distribution $P(\mathbf{Z}^l)$ of one layer, but also joint distribution $P(\mathbf{Z}^1, ldots, \mathbf{Z}^l)$ of several layers. A proper trade-off of the two discrepancies can advance transferability between two domains.

## 2 Related Work

Psychological researches show that human generates different emotions according to different visual content [9,11]. And because of the development of social networks, more and more people upload their images, which increase the image amount for researches. Therefore, emotion researchers pay more attention from the psychology analysis to the image emotion analysis. Some research works even extent the analysis from dominant emotion to personalized emotion [32,34,35]. Traditional method classifying emotion contained in images based on low-level crafted features [8,17,19,37]. For example, Machajdik et al. [18] designed 8 kinds of emotion-related features. Zhao et al. [31] proposed principles-or-art based features for discriminating emotions.

Recently, deep learning has made great development [6,13] and convolutional neural networks (CNNs) are widely applied in computer vision [5,10,22]. One import reason is that the appearance of large-scale datasets, such as ImageNet [1], boosts the features learning of CNNs. In visual emotion analysis, You et al. [28] utilized weakly labeled images to train a CNN and learned a binary image emotion classifier. Then they built a large-scale dataset for image emotion analysis [29]. And the CNN based emotion classifier outperformed ones based manually crafted features [29]. However, training a CNN requires massive labeled data and many emotion-image domains lack them. Although some methods were designed to ease the problem, such as generating images similar the target domains [33,36], the generating procedure is fussy and the qualities of generated images can not be guaranteed.

In this paper, we aim at alleviating the data scarcity problem with transfer learning. Transfer learning focus on knowledge transfer from the source domain with rich label information to the target domain [26]. Traditional transfer learning methods learn domain-invariant model based on shallow features [2,7,20]. Recent studies have demonstrated that deep models can learn more transferable features between two domains [27]. For example, when a CNN extract features from different image domains, the first-layer features all tend to resemble Gabor filters or color blobs.

However, as CNNs always learn domain-specific features at top layers, distributions of different domains exist relatively large discrepancies at these layers. Therefore, many researchers add specific transfer modules to reduce the discrepancies in a layer-wise way [14,15,24]. These methods promote the effect of deep transfer learning. However, it is necessary to consider the dependencies between layers. Long et al. [16] proposed joint adaptation network, which first considered the joint distribution of all the top fully-connected layers.

## 3 Preliminary

### 3.1 Maximum Mean Discrepancy

Maximum Mean Discrepancy (MMD) is used to judge whether two distributions $P(\mathbf{X}^s)$ and $Q(\mathbf{X}^t)$ are the same [4]. Its hypothesis is $\mathbb{E}_P[f(\mathbf{X}^s)] = \mathbb{E}_Q[f(\mathbf{X}^s)]$ when $P = Q$. Now it is usually used to measure the distribution similarity and its form is presented as:

$$D(P, Q) \triangleq \sup_{f \in \mathcal{F}} (\mathbb{E}_P[f(\mathbf{X}^s)] - \mathbb{E}_Q[f(\mathbf{X}^t)]) \tag{1}$$

where $\mathcal{F}$ is a functional set.

### 3.2 Reproducing kernel Hilbert space

MMD can be represented as the distance in Reproducing kernel Hilbert space [4]. As Euclidean space $\mathcal{V}$ is a finite vector space, Hilbert Space is typically viewed as an infinite function space $\mathcal{H}$ and its orthogonal basis can be denoted as $\{\sqrt{\lambda_i}\psi_i\}_{i=1}^{\infty}$, where $\psi_i$ is the base function in each dimension. If $\mathbf{X}$ is a random variable in domain $\mathbf{\Omega}$, a function $f : \mathbf{\Omega} \to \mathbb{R}$ in $\mathcal{H}$ can be presented as $(f_1, f_2, \ldots)_{\mathcal{H}}^T$. And we define another infinite-dimensional feature map $\phi(\mathbf{x})$ in $\mathcal{H}$ as $(\sqrt{\lambda_1}\psi_1(\mathbf{x}), \sqrt{\lambda_2}\psi_2(\mathbf{x}), \ldots)_{\mathcal{H}}^T$. [3,21] We find that:

$$< f, \phi(\mathbf{x}) >= \sum_{i=1}^{\infty} f_i \sqrt{\lambda_i}\psi_i(\mathbf{x}) = f(\mathbf{x}) \tag{2}$$

This is the reproducing property of $\mathcal{H}$. We denote $\mathbb{E}_P[f(\mathbf{X}^s)]$ and $\mathbb{E}_Q[f(\mathbf{X}^t)]$ as $\mu_{\mathbf{x}}(P)$ and $\mu_{\mathbf{x}}(Q)$ respectively [4]. Now MMD can be presented as:

$$\begin{aligned} D(P, Q) &= \sup_{f \in \mathcal{H}} (\mathbb{E}_P[f(\mathbf{X}^s)] - \mathbb{E}_Q[f(\mathbf{X}^t)]) \\ &= \sup_{f \in \mathcal{H}} (\mathbb{E}_P[< \phi(\mathbf{X}^s), f >] - \mathbb{E}_Q[< \phi(\mathbf{X}^t), f >]) \\ &= \sup_{f \in \mathcal{H}} < \mu_{\mathbf{x}}(P) - \mu_{\mathbf{x}}(Q), f > \end{aligned} \tag{3}$$

If we only select $f$ which satisfies $|f| = 1$, $D(P, Q)$ can be calculated as:

$$\begin{aligned} D(P, Q) &= ||\mu_{\mathbf{x}}(P) - \mu_{\mathbf{x}}(Q)||^2_{\mathcal{H}} \\ &=< \mu_{\mathbf{x}}(P), \mu_{\mathbf{x}}(P) > \\ &\quad + < \mu_{\mathbf{x}}(Q), \mu_{\mathbf{x}}(Q) > -2 < \mu_{\mathbf{x}}(P), \mu_{\mathbf{x}}(Q) > \end{aligned} \tag{4}$$

Now we can define a kernel function $k(\mathbf{x}, \mathbf{y})$ to replace $< \phi(\mathbf{x}), \phi(\mathbf{y}) >$. The kernel function can not only be a scalar product, but also other choices like Gaussian kernel. This method is widely employed in many tasks like density estimation and two-sample test [4,23]. Given two sets $\mathcal{D}_s = \{\mathbf{x}_i^s\}_{i=1}^{n_s}$ and $\mathcal{D}_t = \{\mathbf{x}_i^t\}_{i=1}^{n_t}$ with finite instances sampled from $P$ and $Q$. The kernel embeddings are calculated by:

$$\mu_{\mathbf{x}}(P) = \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{x}_i^s) \tag{5}$$

$$\mu_{\mathbf{x}}(Q) = \frac{1}{n_t} \sum_{i=1}^{n_t} \phi(\mathbf{x}_i^t) \tag{6}$$

As $k(\mathbf{x}, \cdot) = \phi(\mathbf{x})$, $\mu_{\mathbf{x}}(P)$ and $\mu_{\mathbf{x}}(Q)$ are called kernel embedding here. Now MMD can be estimated as the distance of two kernel embeddings and its formula is:

$$D(P, Q) = \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} k(\mathbf{x}_i^s, \mathbf{x}_j^s) + \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} k(\mathbf{x}_i^t, \mathbf{x}_j^t) - \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} k(\mathbf{x}_i^s, \mathbf{x}_j^t) \tag{7}$$

## 4 Transfer Learning for Image Emotion Analysis

Given a source emotion-image domain $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{n_s}$ and a target emotion-image domain $\mathcal{D}_t = \{(\mathbf{x}_i^t, y_i^t)\}_{i=1}^{n_t}$, where $n_s \gg n_t$, our task is employing a transfer learning method to optimize a CNN with $\mathcal{D}_s$ and $\mathcal{D}_t$ and improve its classification performance in $\mathcal{D}_t$. The specific method is to reduce the domain distribution discrepancy at the fully-connected layers while training the CNN with $\mathcal{D}_s$ and $\mathcal{D}_t$ simultaneously.

Choosing a CNN as our base transfer learning model is based on two reasons: (1) Compared with conventional manually crafted features, features extracted by CNNs are more suitable for image emotion analysis; (2) Recent studies show that CNNs can learn more transferable image features at first layers.

where $J$ is a cross-entropy loss function. Intuitively, if we hope to utilize $\mathcal{D}_s$ to improve the performance of a CNN on $\mathcal{D}_t$, we can employ both $\mathcal{D}_s$ and $\mathcal{D}_t$ to train the same CNN together. However, in the image emotion analysis field, there always exists a discrepancy between domain distributions $P(\mathbf{X}^s)$ and $Q(\mathbf{X}^t)$. Meanwhile, the image features transits from general to domain-specific along a CNN, which means the transferability decreases at the fully-connected (FC) layers. Our transfer learning method minimizes the domain shift at FC layers from two perspectives: (1) Reducing marginal distribution discrepancy $\{P(\mathbf{Z}^{si}, Q(\mathbf{Z}^{ti})\}_{i \in \mathcal{G}}$ in a layer-wise way; (2) Reducing joint distribution discrepancy $P(\mathbf{Z}^{s1}, \ldots, \mathbf{Z}^{s|\mathcal{G}|})$ and $P(\mathbf{Z}^{t1}, \ldots, \mathbf{Z}^{s|\mathcal{G}|})$. $\{\mathbf{Z}^{si}\}_{i \in \mathcal{G}}$ and $\{\mathbf{Z}^{ti}\}_{i \in \mathcal{G}}$ are features at FC layers. $\mathcal{G}$ is a set of selected fully-connected layers to be aligned for joint distribution. Usually, $\mathcal{G}$ contains all the fully-connected layers of the CNN.

### 4.1 Joint Maximum Mean Discrepancy

To decrease joint distribution discrepancy of two domains, Long et al. [16] designed a module to measure joint distribution discrepancy like MMD, which is called Joint Maximum Mean Discrepancy (JMMD). JMMD is estimated as:

$$D_{\mathcal{G}}(P, Q) \triangleq ||\mathcal{C}_{\mathbf{Z}^s,1:|\mathcal{G}|}(P) - \mathcal{C}_{\mathbf{Z}^t,1:|\mathcal{G}|}(Q)||^2 \tag{8}$$

$\mathcal{C}_{\mathbf{Z}^s,1:|\mathcal{G}|}(P)$ and $\mathcal{C}_{\mathbf{Z}^t,1:|\mathcal{G}|}(Q)$ is the feature embedding in Hilbert space.

$$\mathcal{C}_{\mathbf{Z}^*,1:|\mathcal{G}|} = \frac{1}{n_*}\sum_{i=1}^{n_*} \otimes_{l=1}^{\mathcal{G}} \phi^l(\mathbf{x}_i^l) \tag{9}$$

Where $* \in \{s, t\}$. If we make use of kernel trick, $D_{\mathcal{G}}(P, Q)$ can be estimated as:

$$\begin{aligned} D_{\mathcal{G}}(P, Q) \triangleq & \frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s}\prod_{l\in\mathcal{G}} k^l(z_i^{sl}, z_j^{sl}) \\ & + \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\prod_{l\in\mathcal{G}} k^l(z_i^{tl}, z_j^{tl}) \\ & - \frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j=1}^{n_t}\prod_{l\in\mathcal{G}} k^l(z_i^{sl}, z_j^{tl}) \end{aligned} \tag{10}$$

### 4.2 Deep Transfer Learning Model

We integrate both MMD and JMMD into the FC layers of the CNN, where MMD is used for measuring marginal discrepancy and JMMD is used for measuring joint discrepancy for two domain. The optimizing process is minimizing MMD and JMMD of fully-connected layers while fine-tuning CNN with $\mathcal{D}_s$ and $\mathcal{D}_t$. The loss function is as follows:

$$\mathcal{L} = \mathcal{L}_s + \mathcal{L}_t + \lambda D_{\mathcal{G}}(P, Q) + \eta \sum_{i\in\mathcal{G}} D_i(P, Q) \tag{11}$$

where $\mathcal{L}_s$ and $\mathcal{L}_t$ are classification loss functions for $\mathcal{D}_s$ and $\mathcal{D}_t$ and they are presented as:

$$\mathcal{L}_s = \frac{1}{n_s}\sum_{i=1}^{n_s} J(f(\mathbf{x}_i^s), y_i^s) \tag{12}$$

$$\mathcal{L}_t = \frac{1}{n_s}\sum_{i=1}^{n_t} J(f(\mathbf{x}_i^t), y_i^t) \tag{13}$$

$D_i(P, Q)$ is the MMD loss at $i$-th FC layer. $\lambda$ and $\eta$ are two trade-off parameters. The overall architecture of JAN is shown in Fig. 1.

## 5 Experiment

Experiments focus on the image emotion classification problem. And the purpose is to evaluate whether our transfer learning method can generalize a CNN trained in a large-scale emotion-image domain to another small-scale one better.

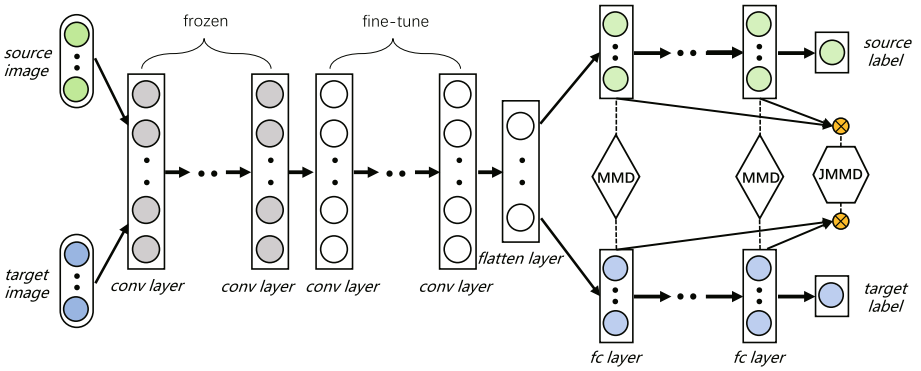


**Fig. 1** Overall architecture of our deep transfer learning model

**Table 1** Statistcs of three existing emotion-image datasets

| Dataset | Amusement | Anger | Awe | Contentment | Disgust | Excitement | Fear | Sadness | Sum |
|---|---|---|---|---|---|---|---|---|---|
| FI | 4861 | 1236 | 3055 | 5292 | 1616 | 2827 | 998 | 2815 | 22,700 |
| ArtPhoto | 101 | 77 | 102 | 70 | 70 | 105 | 115 | 166 | 806 |
| IAPS-Subset | 37 | 8 | 54 | 53 | 74 | 55 | 42 | 62 | 395 |

**Datasets**

**FI** [29] contains 22700 emotion-images in 8 categories. Images are collected through search engines (Flickr and Instagram) with 8 emotion keywords. Then images are labeled using Amazon Mechanical Turk (AMT).

**ArtPhoto** [18] consists of 806 photos from professional artists. The labels of photos are provided by image owners.

**IAPS-Subset** is a subset of the International Affective Picture System (IAPS) [12]. This dataset and **Artphoto** share the same 8 categories with **FI**

Table 1 shows the statistics of the two datasets. For **ArtPhoto** and **IAPS-Subset**, the table shows that image numbers of each category are imbalanced and the total numbers are both much smaller than that of **FI**. Therefore, we take **FI** as the source domain when it is included in the task.

Based on the 3 datasets, we construct 4 cross emotion-image domain classification tasks: $\mathbf{F} \rightarrow \mathbf{A}$, $\mathbf{F} \rightarrow \mathbf{I}$, $\mathbf{I} \rightarrow \mathbf{A}$, $\mathbf{A} \rightarrow \mathbf{I}$. $\mathbf{F} \rightarrow \mathbf{A}$ means, for example, taking **FA** as the source domain and **ArtPhoto** as the target domain.

We randomly split the target-domain data into training, validation, and test set with fractions 80%, 5%, 15%. We perform a 5-fold Cross Validation to obtain results.

**Architecture**

We choose ResNet50 [1] as the base CNN. A fully-connected layer and a softmax layer are added behind convolutional layers. We fine-tune the whole network in an end-to-end way. We measure marginal distribution discrepancy at the FC layer with MMD and joint discrepancy of the FC layer and softmax layer with JMMD. The $\lambda$ and $\eta$ in Eq. 12 are 0.2 and 0.3 respectively.

**Baseline**

– CTD [29]: The CNN model is fine-tuned only with labeled data in target domain. This is the basic method used for image emotion classification.

**Table 2** Accuracy (%) on 4 cross emotion-images classification tasks

| | CTD | CBD | DAN | JAN | Ours |
|---|---|---|---|---|---|
| **F → I** | 24.81 | 26.30 | 25.93 | 27.78 | **29.63** |
| **A → I** | 22.96 | 24.44 | **30.00** | 27.41 | 27.04 |
| **F → A** | **39.66** | 36.92 | 36.24 | 36.75 | 37.61 |
| **I → A** | 30.77 | 34.56 | 36.78 | 34.81 | **39.15** |

The accuracies in bold are highest ones in their corresponding tasks

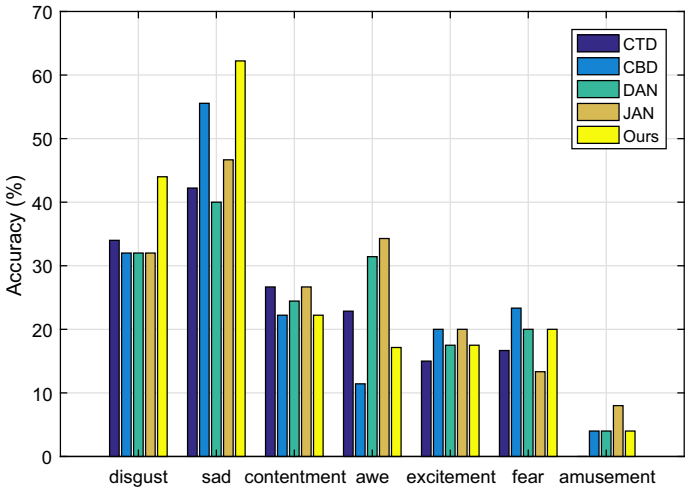


**Fig. 2** Per-emotion accuracy on task **F** → **I**

– CBD: The model is fine-tuned with labeled data in both the source and the target domain without transferring modules.
– DAN [14]: This is a classical deep transfer learning method, which measure the domain distribution discrepancy with MMD and reduce it in a layer-wised way.
– JAN [16]: This method aligns full-conncted layers of a CNN and minimize their joint distribution discrepancy with JMMD.

The CNN architecture and fine-tuned layers for all the baselines are the same as those of our model. The final classification results on target domains are reported in Table 2.

Table 2 reveals the following observations: (1) CBD outperforms CTD on most tasks, which proves the transferability of CNNs; (2) Deep transfer learning method performs better than CBD. This validates that integrating transfer modules into CNNs can boost it to learn more transferable features; (3) Our method outperforms DAN and JAN in most cases, which demonstrates that reducing marginal and joint distribution discrepancies together can improve the transferability of the CNN further; (4) On task **F** → **A**, CTD performs best, which shows that the degree of domain shift influences the feasibility of transfer learning. When the domain shift is large, the transferred information from the source domain may become noise information.

Figures 2 and 3 show the accuracy of each emotion category on task **F** → **I** and **I** → **A**. We do not report the result of emotion *anger* as its data amount is scant. The results reveal that the CNN classifier with transferring modules consistently outperforms conventional CNN classifier. Furthermore, DAN, JAN outperforms our method on partial categories, which

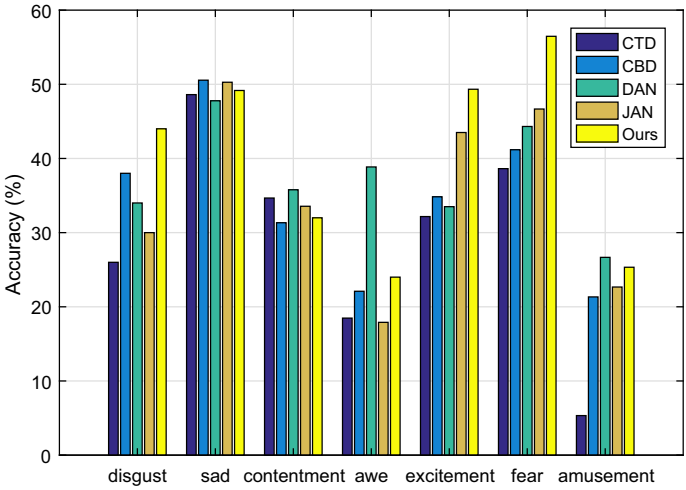


**Fig. 3** Per-emotion accuracy on task $\mathbf{I} \rightarrow \mathbf{A}$

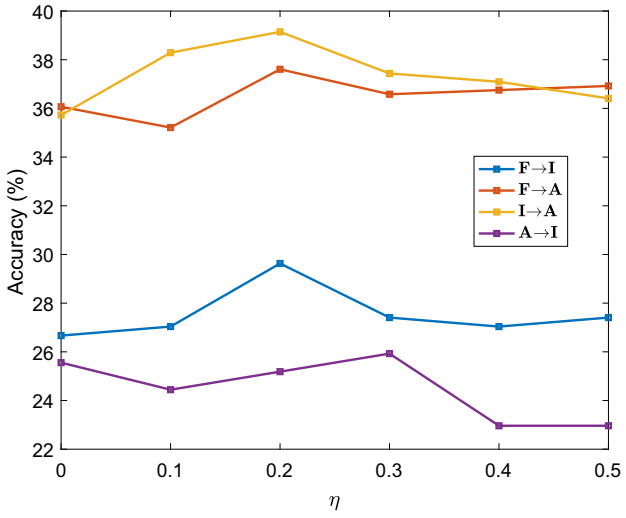


**Fig. 4** Accuracy w.r.t. $\eta$

demonstrates that the most proper ratios between marginal and joint distribution discrepancies are different for different categories. But on most categories, our method performs the best. Therefore, considering the two different discrepancies together is necessary.

**Parameter Analysis**

Now we check the sensitivity of proportion between JMMD parameter $\lambda$ and MMD parameter $\eta$ in Eq. 12. the value of $\eta$ varies in $\{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ and $\lambda = 1 - \eta$. The results are shown in Fig. 4. The results present as bell-shaped curves, which confirms our motivation that a proper trade-off between marginal and joint distribution discrepancies can advance the transferability of CNNs.

### 5.1 Conclusion

In this paper, we propose a deep transfer learning method into image emotion analysis. Our purpose is improving the classification performance of CNNs on a small-scale emotion-image domain by transferring label information from another large-scale one. During the transferring process, we decrease the marginal and the joint distribution discrepancies together. The experimental results demonstrate the promise of our method for discriminating image emotions. In future work, we will explore how to transfer information from art and psychological theory based features to CNN-based features.



## References

1. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: CVPR09
2. Gong B, Grauman K, Sha F (2013) Connecting the dots with landmarks: discriminatively learning domain-invariant features for unsupervised domain adaptation. In: ICML
3. Gretton A (2013) Introduction to RKHS, and some simple kernel algorithms. In: Adv Top Mach Learn Lecture Conducted from University College London 16
4. Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, Smola A (2012) A kernel two-sample test. J Mach Learn Res 13:723–773
5. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR) pp 770–778
6. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18:1527–1554
7. Huang J, Smola AJ, Gretton A, Borgwardt KM, Schölkopf B (2006) Correcting sample selection bias by unlabeled data. In: NIPS
8. Jia J, Wu S, Wang X, Hu P, Cai L, Tang J (2012) Can we understand van gogh's mood?: learning to infer affects from images in social networks. In: ACM multimedia
9. Joshi D, Datta R, Fedorovskaya E, Luong QT, Wang JZ, Li J, Luo J (2011) Aesthetics and emotions in images. IEEE Signal Process Mag 28(5):94–115
10. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
11. Lang PJ (1979) A bio-informational theory of emotional imagery. Psychophysiology 16(6):495–512
12. Lang PJ, Bradley MM, Cuthbert BN (1997) International affective picture system (IAPS): technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention, pp 39–58
13. LeCun Y, Kavukcuoglu K, Farabet C (2010) Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems, pp 253–256
14. Long M, Wang J (2015) Learning transferable features with deep adaptation networks. In: ICML
15. Long M, Wang J, Jordan MI (2016) Unsupervised domain adaptation with residual transfer networks. In: NIPS
16. Long M, Wang J, Jordan MI (2017) Deep transfer learning with joint adaptation networks. In: ICML
17. Lu X, Suryanarayan P, Adams Jr, RB, Li J, Newman MG, Wang JZ (2012) On shape and the computability of emotions. In: Proceedings of the 20th ACM international conference on Multimedia, pp 229–238. ACM
18. Machajdik J, Hanbury A (2010) Affective image classification using features inspired by psychology and art theory. In: ACM multimedia
19. Nicolaou MA, Gunes H, Pantic M (2011) A multi-layer hybrid framework for dimensional emotion classification. In: Proceedings of the 19th ACM international conference on Multimedia, pp 933–936. ACM
20. Pan SJ, Tsang IW, Kwok JT, Yang Q (2009) Domain adaptation via transfer component analysis. IEEE Trans Neural Netw 22:199–210

21. Paulsen VI, Raghupathi M (2016) An introduction to the theory of reproducing Kernel Hilbert spaces, vol 152. Cambridge University Press
22. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. CoRR arXiv:abs/1409.1556
23. Smola AJ, Gretton A, Song L, Schölkopf B (2007) A hilbert space embedding for distributions. In: ALT
24. Tzeng E, Hoffman J, Darrell T, Saenko K (2015) Simultaneous deep transfer across domains and tasks. In: 2015 IEEE international conference on computer vision (ICCV), pp 4068–4076
25. Tzeng E, Hoffman J, Zhang N et al (2014) Deep domain confusion: maximizing for domain invariance. Comput Vis Pattern Recognit. arXiv:1412.3474
26. Weiss KR, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J Big Data 3:1–40
27. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: NIPS
28. You Q, Luo J, Jin H, Yang J (2015) Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: AAAI
29. You Q, Luo J, Jin H, Yang J (2016) Building a large scale dataset for image emotion recognition: The fine print and the benchmark. In: AAAI
30. Zhao S, Ding G, Huang Q, Chua TS, Schuller BW, Keutzer K (2018) Affective image content analysis: a comprehensive survey. In: IJCAI, pp 5534–5541
31. Zhao S, Gao Y, Jiang X, Yao H, Chua TS, Sun X (2014) Exploring principles-of-art features for image emotion recognition. In: ACM multimedia
32. Zhao S, Gholaminejad A, Ding G, Gao Y, Han J, Keutzer K (2019) Personalized emotion recognition by personality-aware high-order learning of physiological signals. In: TOMM
33. Zhao S, Lin C, Xu P, Zhao S, Guo Y, Krishna R, Ding G, Keutze K (2019) Cycleemotiongan: Emotional semantic consistency preserved cyclegan for adapting image emotions. In: AAAI
34. Zhao S, Yao H, Gao Y, Ding G, Chua TS (2016) Predicting personalized image emotion perceptions in social networks. In: IEEE transactions on affective computing
35. Zhao S, Yao H, Gao Y, Ji R, Ding G (2017) Continuous probability distribution prediction of image emotions via multitask shared sparse regression. IEEE Trans Multimed 19(3):632–645
36. Zhao S, Zhao X, Ding G, Keutzer K (2018) Emotiongan: unsupervised domain adaptation for learning discrete probability distributions of image emotions. In: 2018 ACM multimedia conference on multimedia conference, pp 1319–1327. ACM
37. Zhou B, Lapedriza À, Xiao J, Torralba A, Oliva A (2014) Learning deep features for scene recognition using places database. In: NIPS