Neurocomputing 291 (2018) 187–194

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Emotion recognition by assisted learning with convolutional neural networks

Xuanyu He, Wei Zhang*

School of Control Science and Engineering, Shandong University, China

ARTICLE INFO

Article history:
Received 18 February 2017
Revised 15 January 2018
Accepted 20 February 2018
Available online 26 February 2018

Communicated by Dr Xiaoming Liu

Keywords:
Image emotion recognition
Convolutional neural network
Assisted learning
Classification

ABSTRACT

Image emotion is the emotion hidden in or passed by a particular image. In this paper, a novel convolutional neural network is proposed to predict the emotion from an image. The proposed model consists of two parts: a binary positive-or-negative emotion classification network and a deep network for specific emotion recognition. During the network training, an assisted learning strategy is introduced to boost the recognition performance. Experimental results demonstrate that the proposed network is capable of extracting active level features and achieves significant gains in emotion recognition accuracy.

1. Introduction

A photograph or a painting is not only a picture carrying visual information about details of a special moment, but also an artistic conception that the author was about to express. The most important element to be captured in the image is the emotion, which may be the key connection between the author and the viewer. From utter happiness to wrenching heartbreak, images have the power to make human beings feel a full range of emotions. That is because images can cause the former memory and then engender the understanding of visual content. The specific mechanism through which images evoke human emotions is a rich field of research. Some studies [1,2] found that emotional response to an image may depend upon several dimensions, such as composition, colorfulness, spatial organization, emphasis, motion, depth, and presence of humans.

Unlike psychological researches, most computer vision works are trying to predict the emotional reaction on a particular image. This task is about high-level semantics inference of images, which attempts to infer the content of an image and associate high-level semantics to it. The main difficulty is to bridge the gap between low-level features extracted from images and high-level semantics concepts, being emotions here. To be able to recognize the different emotions, features should be designed to carry sufficient information. Some attempts have been made to extract features for emotion recognition in the past years. Dellandrea et al. [3] tried to extract color features and line features from images for affective semantics classification. Machajdik and Hanbury [4] extracted color, texture, composition and content features to represent the emotional content of images. Zhao et al. [5] designed a series of principle-of-art features, which are balance, emphasis, harmony, variety, gradation and movement, to classify and score the emotion of an image. These handcrafted features are all based on the research about how human beings understand the image emotionally.

Inspired by the successes of deep learning in vision, You et al. [6] employed convolutional neural networks (CNNs) to extract features instead of using handcrafted features, trying to solve a binary classification problem. With availability of large scale datasets, CNNs are capable of learning the representations of data in multiple abstracting levels. Previous image emotion recognition researches only rely on several small-sized datasets, such as IAPS-Subset [7], ArtPhoto [4] and Abstract Paintings [4]. These datasets are too limited to extract features with good generalization capability by CNNs. You et al. [8] built a large scale dataset to boost the research in emotion recognition. As shown in Fig. 2, following previous ones, the dataset includes eight kinds of visual emotions, which are *Amusement, Awe, Contentment, Excitement, Anger, Disgust, Fear* and *Sadness*.

In this work, we are interested in using CNNs to extract affective level features and classify the emotions evoked from a series

* Corresponding author.
E-mail address: davidzhang@sdu.edu.cn (W. Zhang).

https://doi.org/10.1016/j.neucom.2018.02.073
0925-2312/

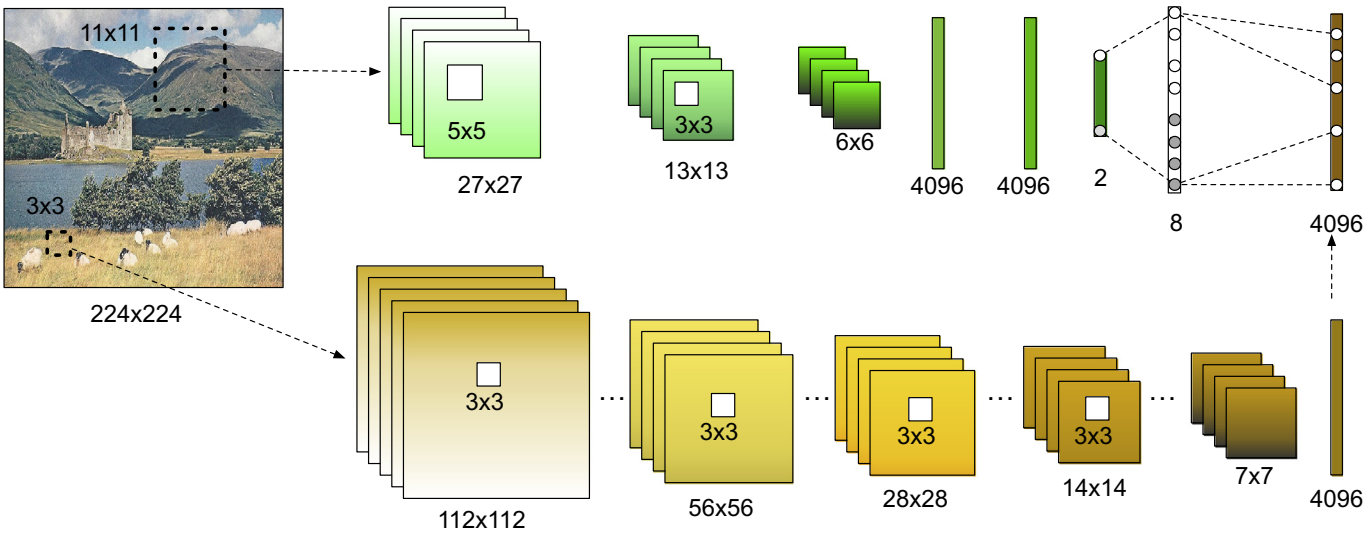


Fig. 1. Structure of our proposed network. Top: B-CNN, a binary positive-or-negative classification network. Bottom: E-CNN, a VGG-16 network for specific emotion recognition. Due to the space limit, the other convolutional layers in the VGG-16 network which have the same config are omitted in this figure.

Fig. 2. Sample images selected from the dataset [8]. The words beneath are the image emotion labels corresponding to those images.

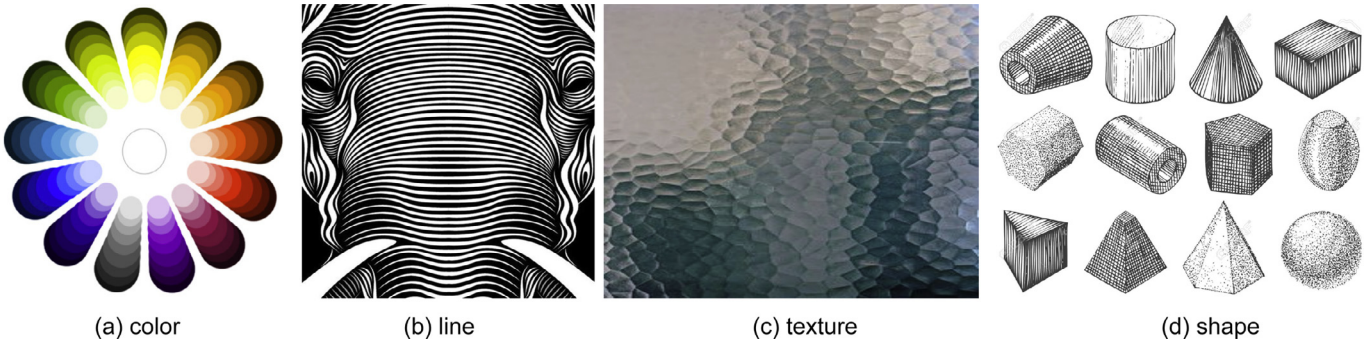


Fig. 3. Typical low-level representation features.

of particular images. Recently, CNNs have been widely employed to solve challenging visual understanding problems [9,10], such as object classification [11–14], semantic segmentation [15,16], object detection [17,18], and object tracking [19,20]. The contribution to CNNs is abundant and increasing. We propose a novel CNN model for image emotion recognition, which is composed of two parts: a binary positive-or-negative emotion network and a deep network to recognize the specific emotion of an image. Considering that the usual emotions can be divided into positive or negative emotion, we think using a binary network to predict that the image emotion is positive or negative will be helpful for the final recognition. The result of the binary network can be considered as a prior knowledge, just like we all know that sadness could not be a positive emotion.

The rest of this paper is organized as follows: we will first briefly review the most relevant literature in Section 2, and then explain our proposed network in detail in Section 3. Extensive experiments are conducted in Section 4 to validate our approach on public emotion dataset. We draw the conclusions of this work in Section 5.

2. Related work

Existing methods of emotion representation can be divided into two types: dimensional approach [21–23] and categorical approach [4,5,24,25]. The dimensional approach allows to represent a wide range of emotions and maps emotions into a two-dimension space, by using few emotion dimensions such as valence and arousal to reveal the main characteristics of emotions. The other approach is categorical approach, consisting of a few basic emotion labels, such as happiness, sadness, fear and anger. Although this kind of method is limited and cannot describe the whole range of emotions, it is a natural way to describe emotion as human beings do in everyday life. The manner of the emotion representation depends on the application. Following most image emotion recognition work, we focus on investigating the categorical approach.

As image emotion recognition is an new research topic, less work was presented to design efficient image features for this task. Traditional image features are still used for image emotion recognition, including color, line, texture and shape, as shown in Fig. 3. Studies on artistic paintings [26,27] have brought about affective level semantics of basic features. Valdez and Mehrabian [28] found a way to map low-level color primitives into emotions. Some colors and their corresponding emotions are illustrated in Table 1. Oblique lines, horizontal lines and vertical lines have different influences on the emotional reaction [3]. Other basic visual features can also carry emotional information.

However, basic features are not invariant to their different arrangements. And humans cannot understand the meanings of these features. According to basic visual features, balance [5], harmony [3,5], variety [5], content [4] and some other features are designed to evaluate the emotional level of images.

Table 1
Human emotions evoked by different colors.

Color	Emotional effects
Orange	Glory
Brown	Relaxation
Purple	Melancholy and fear
Red	Happiness, dynamism and power
Green	Calmness, hope and relaxation
Blue	Gentleness, fairness, faithfulness and virtue

Recently, some efforts were spent on employing CNNs to extract features for visual emotion analysis [6,8,29–31]. Features extracted by CNNs proved to better than the manually designed features [8,32] as CNNs can learn representations on the samples for specific purposes [33].

For example, Xu et al. proposed a visual sentiment prediction framework with CNNs [30]. A CNN was trained for object classification and then transferred to the problem of sentiment prediction. Chen et al. introduced a visual sentiment concept classification method based on a deep CNN structure [29]. This work used a mid-level representation Adjective Noun Pairs (ANPs) as the labels of the images. By combining the sentimental strength of adjectives and detectability of nouns, the ANPs proved to be useful as statistical cues for emotion recognition in images. Campos et al. explored how CNNs could be specifically applied to the task of visual sentiment prediction by visualizing local patterns of the image associated to its sentiment [31]. All these works have shown that CNNs could lead to improvements on visual emotion analysis. However, those works mainly focused on the binary classification problem, recognizing whether the image is positive or negative. Unlike the above work, You et al. [8] built a large scale dataset for image emotion recognition. Inspired by Machajdik and Hanbury [34] and Lang et al. [4], eight types of emotions obtained from psychological studies were defined as the labels of images. The benchmark recognition results were produced using AlexNet [13].

3. Approach

As illustrated in the Fig. 1, the proposed emotion recognition model contains two CNN architectures: a Binary CNN (B-CNN) and an Eight-class CNN (E-CNN). The details of CNNs are described in Section 3.1. An assisted learning strategy is employed to train the model. At the testing phase, the proposed model is considered as a multi-level feature extractor and emotion classifier, that is, classify the image as positive or negative first and then recognize the specific emotion based on the binary result.

Table 2
Statistics of the current image emotion datasets.

Dataset	Amusement	Anger	Awe	Contentment	Disgust	Excitement	Fear	Sadness	Sum
IAPS-subset [7,34]	37	8	54	63	74	55	42	62	395
ArtPhoto [4]	101	77	102	70	70	105	115	166	806
Abstract paintings [4]	25	3	15	63	18	36	36	32	228
Image emotion dataset [8]	4942	1266	3151	5374	1658	2963	1032	2922	23,308

Table 3
Statistics of the downloaded images from image emotion dataset [8].

Amusement	Anger	Awe	Contentment	Disgust	Excitement	Fear	Sadness	Sum
4847	1228	3036	5268	1631	2808	1009	2771	22,598

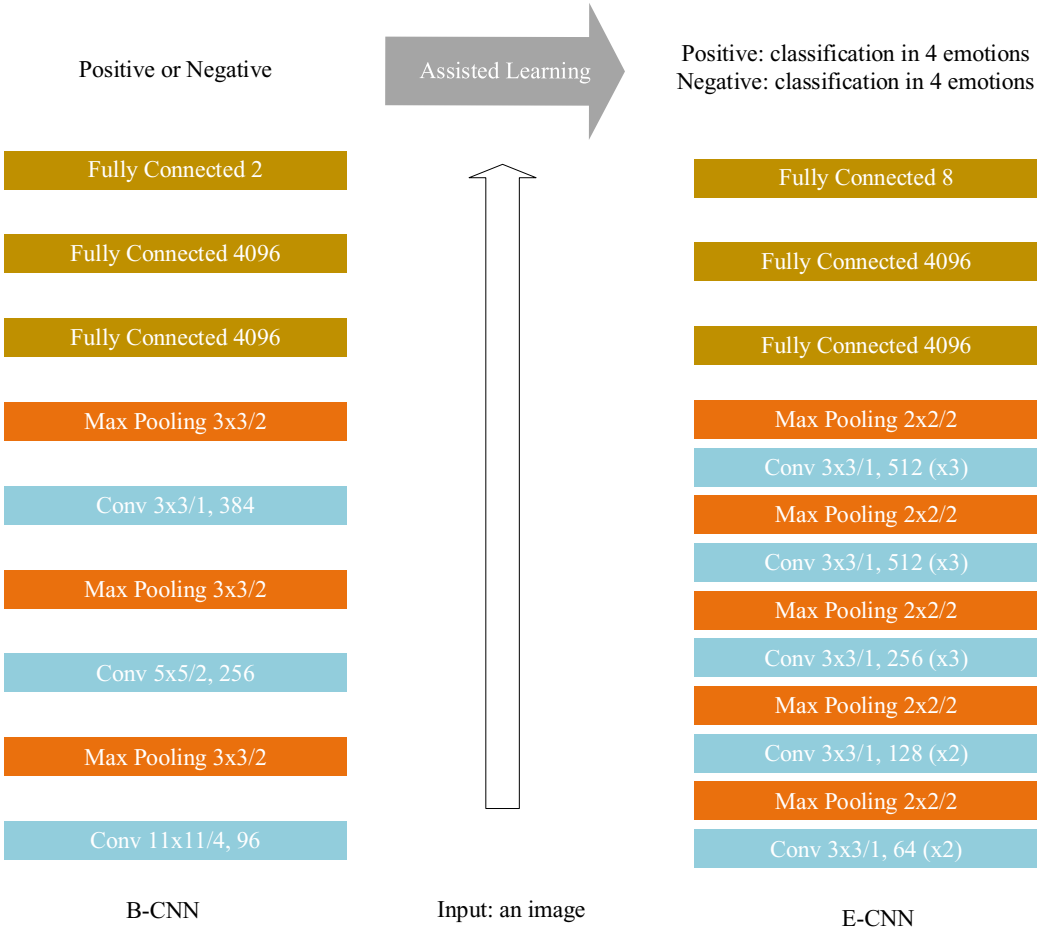


Fig. 4. Details of the proposed model. Layers in E-CNN with the same config are omitted by marking the numbers on the right.

Table 4
Overall accuracy of different networks.

Network architecture	Overall accuracy (%)
AlexNet [13]	32.1
AlexNet (fine-tuned) [8]	58.3
VGG-16 Net (fine-tuned) [12]	61.7
VGG-16 Net with the proposed learning strategy	63.8
Proposed whole network	64.6

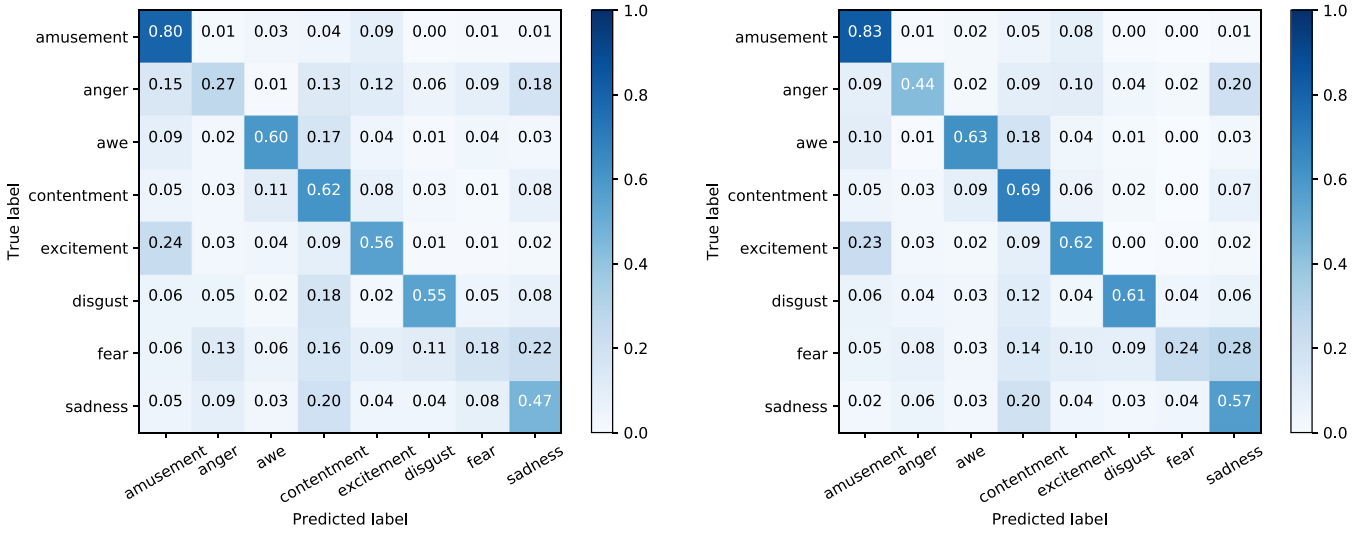
3.1. Model structure

As aforementioned, we first design a CNN model to implement the binary positive-or-negative classification, which is named B-CNN here. It is believed that the binary classification, which is judging that a particular image will pull the viewer into a positive emotional scene or a negative one, is relatively easier and helpful for the final specific emotion recognitionã This proposed CNN contains three convolutional layers and three fully-connected layers. The first convolutional layer filters the input 224×224 RGB image with 96 kernels of size 11×11 with a stride of 4 pixels. The second convolutional layer has 256 kernels of size 5×5 with a stride of two pixels. The third convolutional layer has 384 kernels of size 3×3 with a stride of one pixel. Each convolutional layer is also followed by a max-pooling layer and a normalization layer. The following two fully-connected layers have 4096 neurons each. The last fully-connected two neuron layer is followed by a softmax layer to produce the positive or negative prediction of image emotion (Fig. 4).

Then, another CNN is designed to do the specific classification based on the binary classification result, which is named E-CNN

Table 5
Image emotion classification on testing set.

Emotions	Amusement	Awe	Contentment	Excitement	Positive
Sample size	694	478	794	431	2397
Accuracy(%)	83.0	62.8	68.8	61.7	70.4
Emotions	Anger	Disgust	Fear	Sadness	Negative
Sample size	181	252	155	405	993
Accuracy(%)	43.6	60.7	23.9	57.3	50.5



(a) VGG-16 Net (b) Proposed Network

Fig. 5. Confusion matrices on the testing set of image emotion dataset.

Table 6
Prediction accuracy of different methods on the VSO dataset.

Method	Accuracy(%)
SentiBank [25]	67.7
Xu et al. [30]	70.4
Ours	81.7

here. Focusing on details of each pixel, a convolutional kernel of size 3×3 [12] with a stride of one pixel throughout the whole convolutional layers is adopted in the experiments. Considering different depths of CNNs may lead to different performance on classification, we have tried different network architectures by testing the overall accuracy after 20,000 training iterations, and finally choose a 16 layers network [12] as the main classification part of the whole network. The last fully-connected layer of the proposed network has eight neurons, representing the eight types of emotions studied in this paper.

Finally, these two networks are combined into one network as outlined in Fig. 1. The whole model is trained with an assisted learning strategy.

3.2. Learning strategy

The key idea is straightforward: an image which belongs to positive emotion should not be recognized as negative one. Therefore, the proposed learning strategy proceeds as follows. First, we train the B-CNN to classify images into two categories: positive emotion or negative emotion. To train the binary emotion classifier, we relabeled all the images with the positive or negative labels. With all the weights of the B-CNN fixed, the whole network is trained on the original training set. Then the E-CNN will recognize the image emotion based on the positive-or-negative classification result of the B-CNN. In this step, only a half of neurons in the last fully-connected layer of the E-CNN will be activated. To be more specific, for positive emotions, the neruons representing *Anger, Disgust, Fear* and *Sadness* are inactivated and vice versa. After a certain number of iterations, we make the B-CNN learn optimize its weights again so that the binary classification result can help improve the emotion recognition performance. Hence, such learning strategy is referred as assisted learning in this work.

4. Experimental results

4.1. Dataset

There are some small datasets and a comparatively large scale dataset for image emotion recognition research, all using the same 8 emotion categories (*Amusement, Anger, Awe, Contentment, Disgust, Excitement, Fear* and *Sadness*) as the ground truth of the images. Statistics of these existing datasets are shown in the Table 2.

IAPS-Subset. It is a subset of the International Affective Picture System (IAPS) [34], which has been widely used in affective image classification. Among all IAPS images, 395 images are labeled with the above mentioned eight discrete emotion categories by computing the arousal and valence values of these images [7].

ArtPhoto. This dataset is obtained by using the emotion categories as search terms in the art sharing site [4]. The emotion categories was determined by the artist who uploaded the photo.

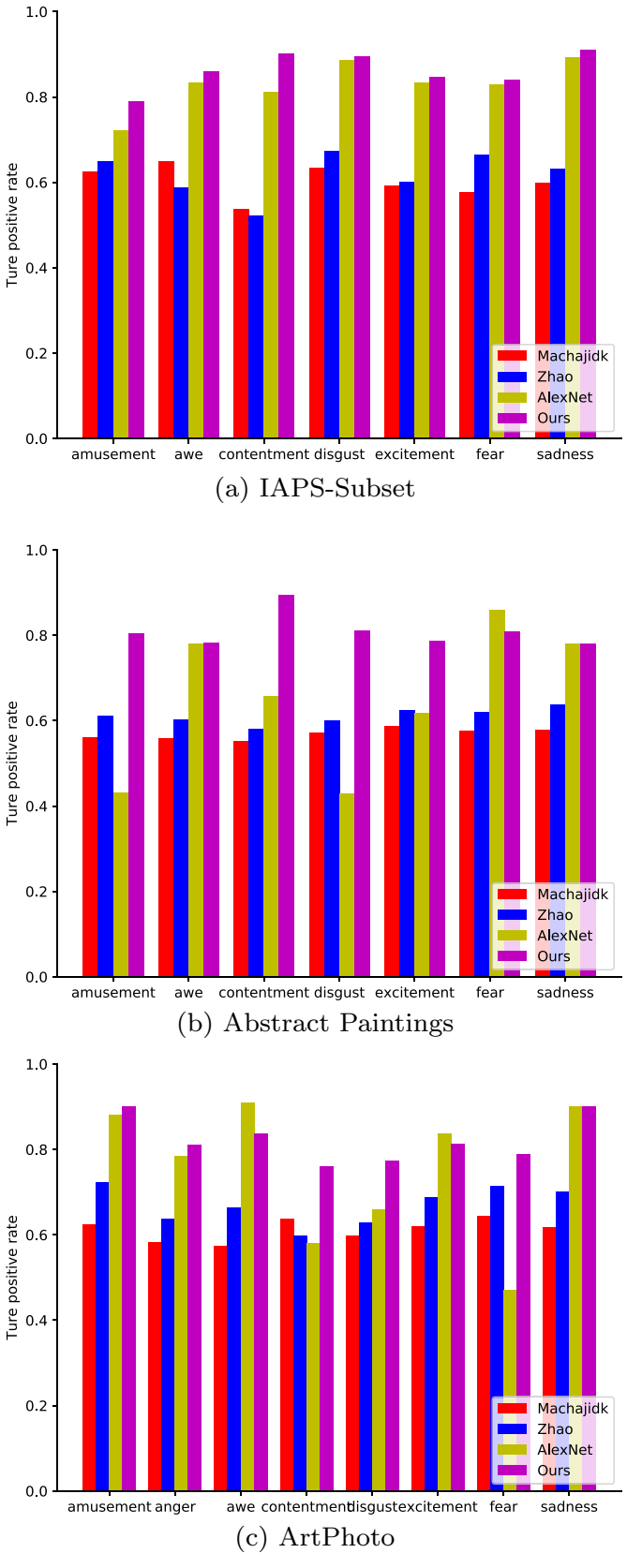


Fig. 6. Emotion recognition comparison on IAPS-Subset, Abstract Paintings and ArtPhoto of the methods such as Machajidk et al. [4], Zhao et al. [5], AlexNet [13] and the proposed model.

These photo are taken by professional artists who attempt to evoke a certain emotion in the viewer of the photograph.

Abstract Paintings. To obtain ground truth for this dataset, the images were peer rated in a web-survey where the participants could select the best fitting emotional category from the ones above [4]. For each image the category with the most votes was selected as the ground truth.

Image emotion dataset. Totally 3 million images are first downloaded from Flickr and Instagram by using the eight emotions above as the keywords for searching [8]. Only 90,000 images are selected to be further labeled by Amazon Mechanical Turk (AMT). AMT workers who passed the qualification test are asked to verify the emotion labels of the images. Each image is verified by five different AMT workers. In total, there are 23,308 strongly labeled images[1].

We use the Image Emotion Dataset to train the proposed deep model. We get 22,598 images of this dataset because some images are unavailable on the Internet now. Table 3 shows the statistics of the remaining images. We randomly selected 85% of images from the dataset as our training dataset, while the remaining 15% samples were used for testing.

4.2. Results

We implement our network using the deep learning framework Caffe [35] on a Linux server with a NVIDIA K40 GPU. We evaluate the performance of our network after training.

4.2.1. Baselines

As a baseline, the AlexNet [8,13] trained on the image emotion dataset lead to an overall accuracy of 32.1%. The overall accuracy of a fine-tuned network [8] is 58.3%. These two networks proved to have better performance on the aforementioned small-sized datasets [4,7] than the methods based on handcrafted visual features [4,5,24,36].

4.2.2. Overall accuracy

Our network has led to an overall accuracy of 64.6% on a randomly selected testing set. To prove the effectiveness of our designs, we take a test on a fine-tuned VGG-16 network [12], which is the main classification part of our proposed network, and yielded an overall accuracy of 61.7%. After assisted learning with the help of the binary classification part, the single VGG-16 network produced an overall accuracy of 63.8%. Table 4 proves our learning strategy is effective to learn more knowledge from images for image emotion recognition.

4.2.3. Recognition on each emotion

Table 5 shows the accuracy of each emotion Fig. 4 on the 15% randomly selected testing images. It is observed in Table 5 that the accuracy of positive emotions are higher than the negative ones. This may be due to the imbalance of positive and negative emotions as shown in Table 3. From the psychological point, it may be also because humans are much easier to feel positive emotion than the negative one.

We also calculate the confusion matrices from the evaluation results on the testing data as shown in Fig. 5. Compared to the baseline VGG-16 network, the proposed model works better at avoid mistaking the positive emotion for negative one. For example, as shown in Fig. 5(a), most *Anger* samples were classified as positive ones (*Amusement*, *Contentment*, and *Excitement*), and only 27% were classified correctly by VGG-16. In contrast, the recognition accuracy benefits from the proposed approach by reducing the

[1] The dataset can be downloaded here: http://www.cs.rochester.edu/u/qyou/deepemotion/index.html.

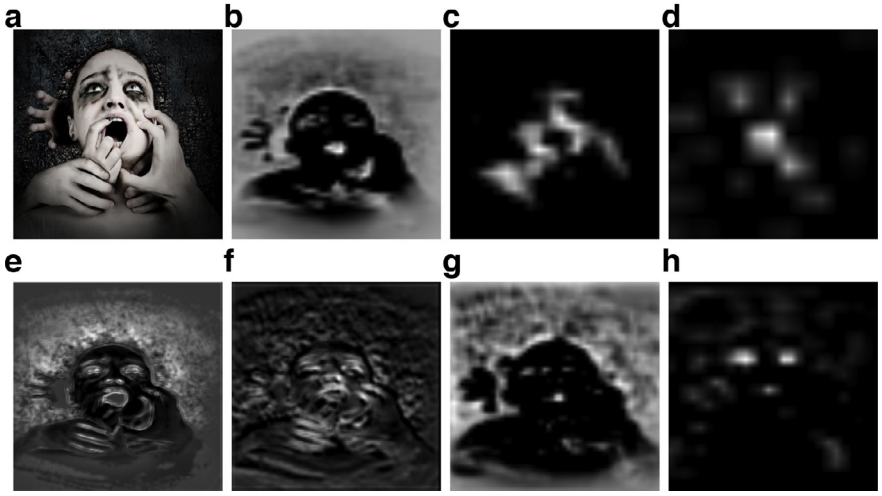


Fig. 7. Process of features extraction.

occurrence of the above situation as shown in Fig. 5(b). It is observed that the recognition accuracy of *Anger* raised by about 17% (from 27% to 44%), and the accuracy of the other categories increased as well.

4.2.4. Recognition on IAPS-Subset, ArtPhoto and Abstract Paintings

To further evaluate the effectiveness of the proposed model, we test the proposed method on the other benchmark datasets such as IAPS-Subset, ArtPhoto and Abstract Paintings, Similar to [8], the image samples from each category were randomly split into five batches and 5-fold cross validation was used to produce the final result. As the emotion category of *anger* only contains 8 and 3 samples in the IAPS-Subset and Abstract Paintings datasets respectively, it is insufficient to perform the 5-fold cross validation. Thus, the true positive rates for emotion *anger* on these two datasets are not reported, following the previous work [4,8].

Similarly, we compare the performance of the proposed model with existing methods, including the hand-crafted methods like Machajdik et al. [4] and Zhao et al. [5], and the deep model like AlexNet [13]. Fig. 6(a)–(c) show the performance of different methods on the three datasets respectively. It can be concluded that the deep models show superiority over the hand-crafted methods in terms of emotion recognition accuracy. Also, the proposed deep model produced the best recognition accuracy for most emotion categories.

4.2.5. Feature visualization analysis

Fig. 7 shows the process of feature extraction from an image with our proposed model. The sample (a) is a horrible image with the emotion label *Fear*. Fig. 7(b), (c) and (d) are the outputs of the first, the second and the third convolutional layer of the B-CNN, respectively. The bottom images (e)–(h) come from the first, the third, the fifth and the seventh convolutional layer outputs of the E-CNN, respectively. Apparently, the E-CNN keeps more details of the image, while the B-CNN seems more abstract.

It seems that the binary positive-or-negative classification part may focus on the global features of images, whereas the main classification part focuses more on the local features of images. We think it is because the two parts use different sizes of convolutional kernels and their purposes are of different levels. Emotions described like *Sadness* and *Contentment* are more detailed than the general *Positive* and *Negative*. So the main classification part needs to extract more details than the binary classification part for specific emotion recognition.

4.3. Emotion classification on VSO dataset

To further evaluate the generalization of our method, we also conducted emotion recognition tests on the Visual Sentiment Ontology (VSO) dataset [25]. VSO dataset is most used for sentiment prediction of the images, which consists of more than 3000 Adjective Noun Pair (ANP) labels. Followed the experimental practice in [25,30], we compare the proposed method with state-of-the-arts methods on the VSO dataset, including Sentibank [25] and Xu et al. [30]. The results shown in Table 6 demonstrate that even without using the mid-level representations, the proposed method outperformed [25] and [30] in terms of emotion recognition accuracy.

5. Conclusions

In this work, the challenging image emotion recognition was investigated. We proposed a novel CNN and design an assisted learning strategy to boost the application of CNNs to visual emotion recognition. Experimental results on benchmark dataset prove that our model achieves significant gains compared to previous networks. In addition, the proposed learning strategy can further improve the performance of CNNs on emotion recognition.

Acknowledgments

This work was supported by the NSFC Grant no. 61573222, Shenzhen Future Industry Special Fund JCYJ20160331174228600, Major Research Program of Shandong Province 2015ZDXX0801A02, National Key Research and Development Plan of China under Grant 2017YFB1300205 and Fundamental Research Funds of Shandong University 2016JC014.

References

[1] P.J. Lang, M.M. Bradley, B.N. Cuthbert, Emotion, motivation, and anxiety: brain mechanisms and psychophysiology, Biol. Psychiatry 44 (12) (1998) 1248–1263.
[2] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J.Z. Wang, J. Li, J. Luo, Aesthetics and emotions in images, Signal Process. Mag. IEEE 28 (5) (2011) 94–115.
[3] E. Dellandrea, N. Liu, L. Chen, Classification of affective semantics in images based on discrete and dimensional models of emotions, in: Proceedings of the International Workshop on Content-Based Multimedia Indexing (CBMI), IEEE, 2010, pp. 1–6.
[4] J. Machajdik, A. Hanbury, Affective image classification using features inspired by psychology and art theory, in: Proceedings of the international conference on Multimedia, ACM, 2010, pp. 83–92.
[5] S. Zhao, Y. Gao, X. Jiang, H. Yao, T.-S. Chua, X. Sun, Exploring principles-of-art features for image emotion recognition, in: Proceedings of the ACM International Conference on Multimedia, ACM, 2014, pp. 47–56.

[6] Q. You, J. Luo, H. Jin, J. Yang, Robust image sentiment analysis using progressively trained and domain transferred deep networks, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.
[7] J.A. Mikels, B.L. Fredrickson, G.R. Larkin, C.M. Lindberg, S.J. Maglio, P.A. Reuter-Lorenz, Emotional category data on images from the international affective picture system, Behav. Res. Methods 37 (4) (2005) 626–630.
[8] Q. You, J. Luo, H. Jin, J. Yang, Building a large scale dataset for image emotion recognition: the fine print and the benchmark, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI Press, 2016, pp. 308–314.
[9] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: a review, Neurocomputing 187 (2016) 27–48.
[10] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing (2016).
[11] D.C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, J. Schmidhuber, Flexible, high performance convolutional neural networks for image classification, in: Proceedings of the International Joint Conference on Artificial Intelligence, 22, 2011, p. 1237.
[12] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).
[13] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the Advances in neural information processing systems, 2012, pp. 1097–1105.
[14] S. Zhou, Q. Chen, X. Wang, Active deep learning method for semi-supervised sentiment classification, Neurocomputing 120 (2013) 536–546.
[15] A. Kendall, V. Badrinarayanan, R. Cipolla, Bayesian segnet: model uncertainty in deep convolutional encoder-decoder architectures for scene understanding, arXiv preprint arXiv:1511.02680 (2015).
[16] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[17] C. Szegedy, A. Toshev, D. Erhan, Deep neural networks for object detection, in: Proceedings of the Advances in Neural Information Processing Systems, 2013, pp. 2553–2561.
[18] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2015, pp. 91–99.
[19] L. Wang, W. Ouyang, X. Wang, H. Lu, Visual tracking with fully convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3119–3127.
[20] H. Xue, Y. Liu, D. Cai, X. He, Tracking people in rgbd videos using deep learning and motion clues, Neurocomputing 204 (2016) 70–76.
[21] X. Lu, P. Suryanarayan, R.B. Adams Jr, J. Li, M.G. Newman, J.Z. Wang, On shape and the computability of emotions, in: Proceedings of the 20th ACM international conference on Multimedia, ACM, 2012, pp. 229–238.
[22] M.A. Nicolaou, H. Gunes, M. Pantic, A multi-layer hybrid framework for dimensional emotion classification, in: Proceedings of the 19th ACM international conference on Multimedia, ACM, 2011, pp. 933–936.
[23] C.H. Chan, G.J. Jones, Affect-based indexing and retrieval of films, in: Proceedings of the 13th annual ACM international conference on Multimedia, ACM, 2005, pp. 427–430.
[24] W.-n. Wang, Y.-l. Yu, S.-m. Jiang, Image retrieval by emotional semantics: a study of emotional space and feature extraction, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics., 4, IEEE, 2006, pp. 3534–3539.
[25] D. Borth, R. Ji, T. Chen, T. Breuel, S.-F. Chang, Large-scale visual sentiment ontology and detectors using adjective noun pairs, in: Proceedings of the 21st ACM international conference on Multimedia, ACM, 2013, pp. 223–232.
[26] R. Arnheim, Art and Visual Perception: A Psychology of the Creative Eye, Univ of California Press, 1954.
[27] C. Colombo, A. Del Bimbo, P. Pala, Semantics in visual information retrieval, IEEE MultiMed. (3) (1999) 38–53.
[28] P. Valdez, A. Mehrabian, Effects of color on emotions., J. Exp. Psychol. Gen. 123 (4) (1994) 394.
[29] T. Chen, D. Borth, T. Darrell, S.-F. Chang, Deepsentibank: visual sentiment concept classification with deep convolutional neural networks, arXiv preprint arXiv:1410.8586 (2014).
[30] C. Xu, S. Cetintas, K.-C. Lee, L.-J. Li, Visual sentiment prediction with deep convolutional neural networks, arXiv preprint arXiv:1411.5731 (2014).
[31] V. Campos, B. Jou, X. Giro-i Nieto, From pixels to sentiment: fine-tuning cnns for visual sentiment prediction, Image Vis. Comput. (2017).
[32] Y. LeCun, K. Kavukcuoglu, C. Farabet, et al., Convolutional networks and applications in vision., in: Proceedings of the ISCAS, 2010, pp. 253–256.
[33] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, Y.L. Cun, Learning convolutional feature hierarchies for visual recognition, in: Proceedings of the Advances in Neural Information Processing Systems, 2010, pp. 1090–1098.
[34] P.J. Lang, M.M. Bradley, B.N. Cuthbert, et al., International Affective Picture System (IAPS): Instruction Manual and Affective Ratings, The Center for Research in Psychophysiology, University of Florida, 1999.
[35] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: Proceedings of the ACM International Conference on Multimedia, ACM, 2014, pp. 675–678.
[36] V. Yanulevskaya, J. Van Gemert, K. Roth, A.-K. Herbold, N. Sebe, J.-M. Geusebroek, Emotional valence categorization using holistic image features, in: Proceedings of the 15th IEEE International Conference on Image Processing, IEEE, 2008, pp. 101–104.

Xuanyu He received the B.S. degree from the Zhejiang University in 2014. He is currently pursuing the M.S degree with the School of Control Science and Engineering, Shandong University, China. His research interests include computer vision, image processing and pattern recognition.

Wei Zhang received the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong in 2010. He is currently with the School of Control Science and Engineering, Shandong University, China. He has authored about 40 papers in international journals and refereed conferences. His research interests include computer vision, image processing, pattern recognition, and robotics. He served as a Program Committee Member and a Reviewer for various international conferences and journals in image processing, computer vision, and robotics.