Multimedia Tools and Applications (2020) 79:14341–14355
https://doi.org/10.1007/s11042-018-6436-0

IDEA: A new dataset for image aesthetic scoring

Xin Jin[1,2] · Le Wu[1] · Geng Zhao[1] · Xinghui Zhou[1] · Xiaokun Zhang[1] · Xiaodong Li[1]

Received: 6 April 2018 / Revised: 14 June 2018 / Accepted: 20 July 2018 /
Published online: 18 August 2018

Abstract
The aesthetic quality assessment of image is a challenging work in computer vision field. The recent research work used the deep convolutional neural network to evaluate the aesthetic quality of images. However, the score of image data sets has a strongly normal distribution, which makes the training of neural network easy to be over-fitting. In addition, traditional deep learning methods usually pre-process images, which destroy the original aesthetic features of the picture, so that the network can only learn some superficial aesthetic features. This paper presents a new data set what images distributed evenly for aesthetics (IDEA). This data set has less statistical characteristics, which is helpful for the neural network to learn the deeper features. We propose a new spatial aggregation perception neural network architecture which can control channel weights automatically. Our experiments in different data sets can prove the advantages and effectiveness of our method.

Keywords Aesthetic assessment · Neural network · Computer vision

1 Introduction

Recently, deep convolutional neural network technology has made great progress object recognition and semantic recognition in the field of computer vision. However, to identify or evaluate the images aesthetic quality by using computer is far from practical. Subjective Image Aesthetic Quality Assessment (IAQA) is still a challenging task [27]. The reason lies in: the images with high aesthetic quality and the images with low aesthetic quality have the big differences, a large number of lower and higher aesthetic characteristics, human evaluation of subjectivity and so on. The aesthetic quality evaluation of images is a hot topic in the field of computer vision, computational aesthetics and computational photography.

The traditional image quality evaluation aims at evaluating the distortion of the image automatically by computer simulation of human visual system. It is mainly aimed at the image quality degradation in the process of acquisition, compression, processing,

✉ Xiaodong Li
lxd@besti.edu.cn

[1] Beijing Electronic Science and Technology Institute, Beijing, 100070, China

[2] CETC Big Data Research Institute Co.,Ltd., Guiyang, Guizhou, China

transmission and display, which usually includes distortion caused by poor imaging conditions and distortion caused by lossy compression. Noise and distortion caused by channel attenuation during image transmission. Although the goal of the experiment is to obtain an objective evaluation value which is consistent with the subjective evaluation result, the purpose of image aesthetic quality evaluating is letting the aesthetic ideology to be reflected by the computer. With the help of computers, we can simulate the human's knowing and understanding of image aesthetics, the characteristics of aesthetic thinking are reflected in the model, and finally the computer can fully capably separate the high-quality images and the low-quality images.

Most of the existing technologies only pay attention to only simple good or bad aesthetic classification, seldom predict the score of pictures, and provide a method of image aesthetic quality scoring based on deep convolution neural network. The convolution neural network has unique advantages in the image processing square with the special structure of its local weight sharing, and its layout is more important. Close to the actual biological neural network, it can effectively simulate human perception of aesthetics.

1.1 Related work

As summarized by [2], the early work of image aesthetic quality evaluation mainly focuses on the manual design of various image aesthetic features, and connected with a machine classification or regression method. Another line is to use generic image description features. Recently, the powerful representation with deep feature, learned from a large amount of data, has shown an ever-increased performance on this task, surpassing the capability of conventional hand-crafted features [3, 11–16, 18–21, 26, 31]. Some works achieve cross-modal retrieval using adversarial learning [6, 34]. The training data for image aesthetic quality assessment usually comes from the online professional photography community, such as photo.net and www.dpchallenge.com. People can rate photos on these sites (scoring 1-7 or 1-10). Most of the above research work [22–25, 30] follows the following methods to predict the quality of image aesthetics:

- Image preprocessing: In order to avoid fitting during training, the image is usually treated with some simple processing to enhance the robustness of the model before training. The main methods include color processing, such as brightness, saturation, contrast change, scaling, cropping, scaling, and flip.
- Convolutional neural networks: Using the classical convolutional neural network architecture of the classification task, or designing and transforming the network to extract the features of image aesthetics, calculate the loss and result according to different aesthetic tasks.

Although aesthetic quality evaluation exists in a common sense, it is still an inherently subjective visual task. For the same image, simple changes may make the opposite judgement. Image preprocessing technology in traditional classification task has been very mature, but in the field of aesthetics, this approach may cause people to make the abstract semantic judgement, which can produce different results.

The quality evaluation of image aesthetics is ambiguous [17]. Figure 1 shows the effect of data augmentation on image aesthetic semantics. Obviously, just changing the position of an image can have such a huge impact. In the field of object detection, He et al. [4] proposed the spatial pyramid pooling layer in the neural network to deal with images of different sizes. In the field of aesthetics, some work has begun to use modified or generated evaluation score distribution to train and give the binary classification results of aesthetic

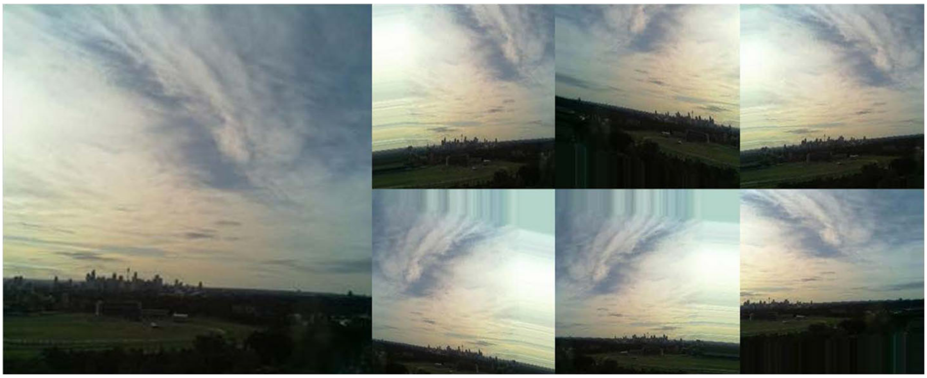


Fig. 1 There are original and preprocessed images. The left image represents the original image, and the right image is produced by the random combination of translation, cropping and left-right flips. Images are from the AVA dataset [29], which contains a list of photo IDs from www.dpchallenge.com

image quality or one-dimensional numerical evaluation [7, 10, 32]. Wu et al. [33] begin to train on small data sets. Mai et al. [27] used adaptive convolution to process the original image, and retained the original aesthetic image features.

On aesthetic data set, Murray et al. [29] the first puts forward a large-scale database for aesthetic visual analysis(AVA). Then, in view of the imbalance of AVA samples, Kong et al. [18] proposed the AADB data set to make the aesthetic data set more balanced which better fit in the normal distribution.

1.2 Our approach

Which images distributed evenly for aesthetics (IDEA) in our data set is presented in our paper. The IDEA data set contains 9191 images, 0-8 points with 1000 images per score, and 9 point has 191 pictures. The obvious distribution features of the data set make the network output keep approaching the bigger weight of the training set's distribution. This paper presents a new spatial aggregation awareness neural network architecture (SAP-Net). We used the squeeze-and-excitation structure to perceptually learn characteristics of different channels proposed by Hu et al. [8]. At the end of the network, the spatial aggregation of these channel characteristics is carried out so that the local information of the network can be merged to form a complete picture aesthetic semantic feature. Behind the network part, we use parallel tasks through divide score on a scale of different particle size and finally formed the global characteristics. The experimental results show that the proposed network and method have better performance in different data sets. The main contributions of this paper are as follows:

- This is the first time to propose a completely balanced aesthetic data set;
- The first work is that channel sensing technology was applied to aesthetic mission, and we put forward a new space fusion strategy to dynamic control channel weighting flow to extract aesthetic characteristics and combine all of local characteristics;
- A multi-task network learning strategy based on fractional granularity is proposed.

In addition to the overall aesthetic quality evaluation of the image, image aesthetics has other goals. This paper predicts that the quality score of image aesthetics can be used for the application of aesthetic image retrieval, photography technical guidance, video cover automatic generation, etc. The evaluation of the quality of image aesthetics has a guiding effect on the application of UAV shooting, robot intelligence, and so on. Only by making the machine with good eyesight can we serve the human beings better.

2 IDEA data set

To create a more balanced set of aesthetic images, we collected several images and score tags from the professional photography website: dpchallenge and Flickr. Website's scores are ranging from 0 to 9, we selected 1000 images for each segment as far as possible (Flickr's label come from AADB).

Specifically, we climbed all the pictures numbered below 800 thousand on dpchallenge which has 330 thousand pictures and we selected 1000 pictures per score. The method of selection is chosen according to the number of voters in the website, because the higher the voting is, the higher the reference value is. Through this method, we get 7961 pictures, mainly missing 1 and 2 points and 9 points, so we randomly select and supplement the images from the AADB data set, and finally form a complete set of data. Finally, our data set which is almost balance distributed, is named as the IDEA dataset. The IDEA data set has 9191 images, of which the number of 9 points number are 191 and the rest are 1,000. In the training, 8191 pictures were used for the training set and randomly selected 1000 images for testing.

Figures 2 and 3 show that the IDEA dataset is more evenly distributed and has less statistical property than the AVA and AADB data sets. We will illustrate the balance of data sets from two aspects. On the first side, think of average and standard deviation, the IDEA

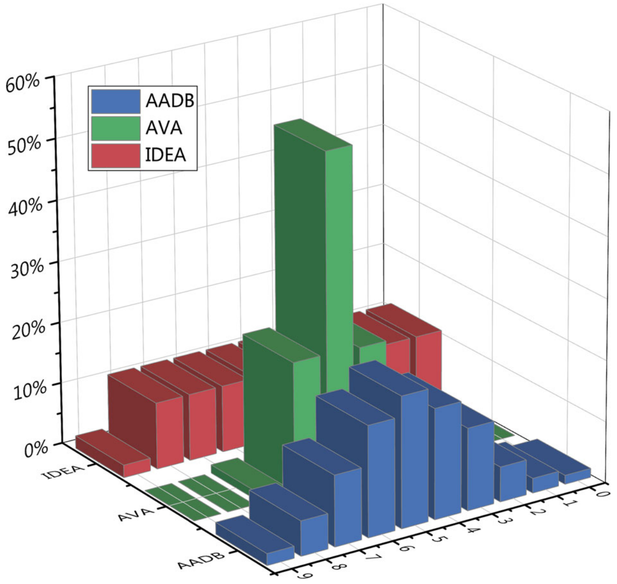


Fig. 2 From left to right: IDEA, AVA, and AADB data sets. X-axis is the fraction, Y-axis is the proportion of numbers. The intermediate fraction of the AVA dataset accounts for the most of total, and the number of images in other segments is seriously unbalanced. The AADB data set is closer to the normal distribution, and the distribution is relatively smooth. The IDEA distribution is almost completely balance

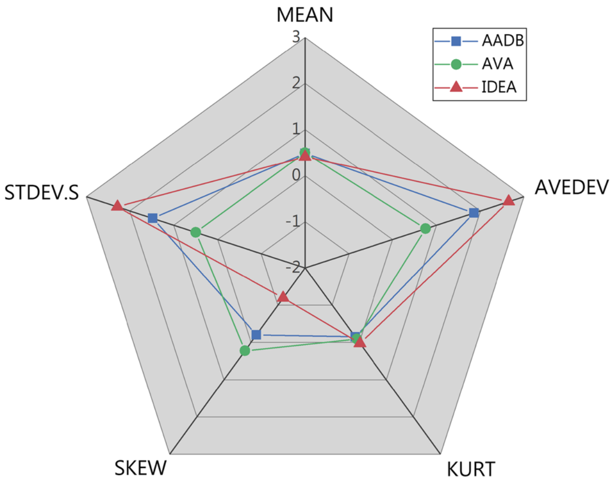


Fig. 3 The relative radar maps of the statistics of AVA, AADB, and IDEA. Counterclockwise from the top represent the mean, absolute mean variance, kurtosis, skewness, standard deviation. The mean, absolute mean deviation, skewness and standard deviation of the three data sets are almost the same standard

data set has the highest mean and standard deviation, reflecting the strongest change in the data set and more extensive distribution, so the trained model has a stronger generalization ability. On the other hand, on the skewness, IDEA has the minimum deviation. The bias is the measure of the skew direction and degree of the distribution of statistical data, and is the digital feature of the asymmetric degree of the distribution of statistical data. This also proves that the dataset has good distribution symmetry and balance characteristics. The AVA data set has the highest skewness due to the serious unbalance of each score segment. The relative mitigation skewness of the AADB dataset is centered. IDEA distribution is more balanced, and skewness is the smallest. However, the normal distribution cannot fit all AVA data, and the aesthetic score cannot simply be regarded as normal distribution.

In addition, the obvious distribution features of AADB make neural network learning only remembers those simple statistical property and shallow aesthetic features, so that the network cannot learn deeper features. In this regard, we propose to remove the statistical property that the neural network can easily remember in the IDEA dataset, and make the network directly learning the semantic features and abstract features of the images so that the performance of the network can be concentrated in the purest aesthetic tasks.

3 Spatial aggregation perception network

3.1 Perception polymerization

In neural networks, the information of different channels in the convolution kernel can be regarded as different perspectives. In this paper, we use the Squeeze-and-Excitation block proposed by Hu et al. [8] to realize channel dimension perception. As shown in Fig. 4, the Squeeze-and-Excitation block module was applied to the ResNet-50 [5] in this article and added to each Residual unit. Here we use global average pooling as the Squeeze operation. The two Fully Connected layers form a Bottleneck structure to model the correlation

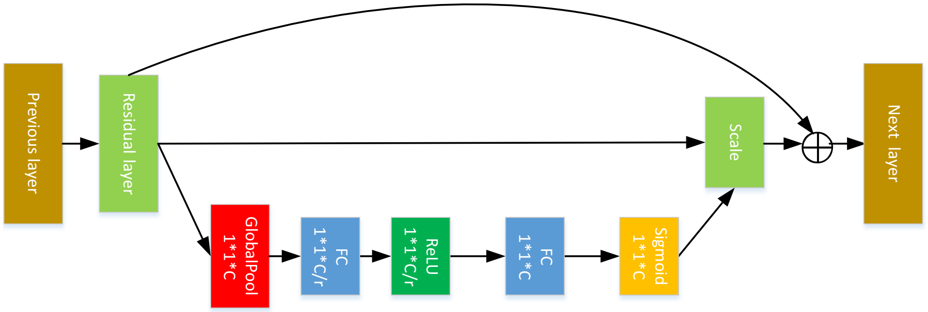


Fig. 4 The Squeeze-and-Excitation block module. Here we use global average pooling as a Squeeze operation. The two Fully Connected layers are then formed to form a Bottleneck structure to model the correlation between the channels and to output and input the same number of weights. We first reduce the feature dimension to the input 1/16, then go through the ReLu activation and then go back to the original dimension through a Fully Connected layer

between the channels and to output the same number of weights as the input features. The function of this module can be expressed by the following formula:

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * x^s \tag{1}$$

Among them, u_c represents the output, C represents the output dimension, C$'$ represents the input dimension. Here is v_c^s a 2D spatial kernel, and therefore represents a single channel of v_c which acts on the corresponding channel of X. Since the output is produced by a summation through all channels, the channel dependencies are implicitly embedded in v_c, but these dependencies are entangled with the spatial correlation captured by the filters.

In traditional neural networks, the output of the convolutional layer does not consider the dependence on each channel. The goal of this paper is to allow the network to selectively enhance the characteristics of large amounts of information, so that subsequent processing can make full use of these features and be useless. Features are suppressed. First, the network encodes global information to achieve compression operations, examines the signal for each channel of the output feature, compresses global spatial information into channel descriptors, and uses global average pooling to generate statistics for each channel. The second is to examine the dependence degree of each channel, and design the adaptive dependence adjustment activation mechanism. There are two criteria for implementing the function: first, to be flexible, and second, to learn a non-mutually exclusive relationship because multiple channels may have results influences. This is achieved using a threshold mechanism with a sigmoid activation function. In order to limit the complexity of the model and enhance the generalization ability, two full connection layers of bottleneck are used in the threshold mechanism. The first FC layer reduces the dimension to 1/r, and r is a hyperparameter. The final sigmoid function is the weight of the each channel. Adjusting the weight of each channel feature which based on the input data can help to enhance the distinguishability of features.

3.2 Spatial aggregation

Traditional neural networks often feature fusion directly through the full connectivity layer, but the relative position of the image cannot be maintained due to the complex combination of the full connectivity layers. In addition, since the number of nodes in the final full connection layer is often fixed in the neural network, which destroys the composition information of the pictures and has an impact on the aesthetic judgment. In response to this problem, He et al. [4] proposed to increase the spatial pyramid pooling layer in neural networks to deal with different size feature maps.

The spatital pyramid pooling layer performs three convolution operations on each of the features of the previous convolutional feature maps. The rightmost image is the original image, the middle one is divided into 9 characteristic features, and the rightmost one is divided into 49 feature images. As shown in Fig. 5, we use a three-layer pyramid pooling layer to set how many blocks the image is divided into. The set of the paper is (1, 3, 7), and then this feature map is processed separately according to the level. That is, the entire feature map of this feature map is pooled at the first level. The maximum pooling is used in the paper and one feature is obtained. The second layer first divides the feature map into 9 small feature maps, then pools them using the corresponding pooling kernels to get 9 features, and the third layer first divides the feature map into 49 pieces. The small feature map is then pooled using pooled kernels of corresponding sizes to obtain 49 features. Then the 1+9+49=59 features are input to the fully connected layer for weight calculation.

3.3 Fractional granularity of the multi-task learning

Multi-task learning is a machine learning method as opposed to single-task learning. In machine learning, the standard algorithmic theory is to learn a task at a time, and the output of the system is a real number. The complex problem in machine learning is first decomposed into theoretically independent sub-problems, and then each sub-problem is studied separately. Finally, the mathematical model of the complex problem is established through the combination of the sub-problem learning results. Multi-tasking learning is a kind of joint learning, in which multiple tasks are studied in parallel and the results influence each other. Multi-task learning means solving multiple problems at the same time.

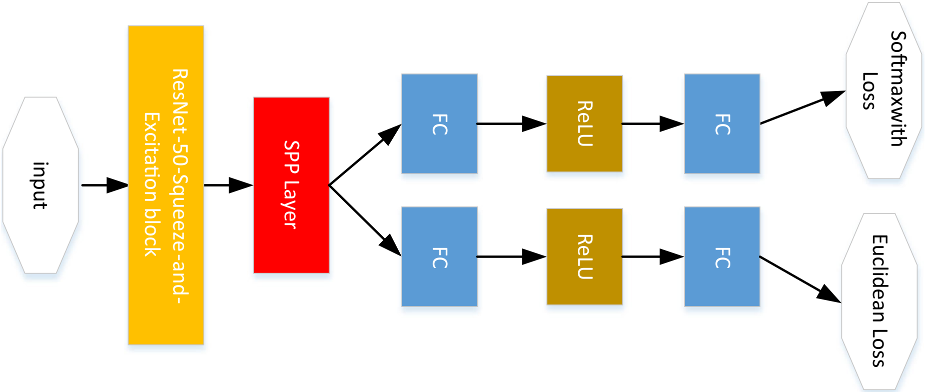


Fig. 5 The architecture of the Multi-task learning. Here we put the last layer of convolutional results into the SPP Layer for spatial aggregation, and then put the aggregation results into different fully-connected structures for spatial information fusion. After a ReLU activation, we output different dimensions of full-connection results according to different tasks. Finally, use different loss calculation formulas for different results to backpropagation the network

Multi-tasking learning is very common in deep learning. The constraints of different tasks can enhance the results of major tasks. Traditional multitasking is often based on the results of generating multiple similar problems through a network, often with some intrinsic links between these issues. In the aesthetic evaluation, the impact of content category factors on classification is not great, and the types of pictures are often more, and the classification network is difficult to converge. From a psychological point of view, it is often difficult for people to give precise and specific scores when viewing a picture given a score. It is difficult for people to give a specific reason for a picture with a small difference, but it is easy for people to tell each specific picture. He is good, medium or bad. Therefore, this feature is applied in this paper and the granularity of classification is discussed.

$$Loss = Loss_{RegLoss} + Loss_{SoftmaxLoss} \tag{2}$$

$$Loss = \frac{1}{2N}\sum_{i=1}^{N}\left\|\hat{y^i} - y^i\right\|_2^2 - \frac{1}{N}\sum_{i=1}^{N} log\frac{e^{f(w,x,b)}}{\sum_{c=1}^{C} f(w,x,b)} \tag{3}$$

The above formula represents the loss calculation formula for multitasking learning used in this paper. The formula consists of two parts, namely, regression loss and classification, that is, Loss$_{RegLoss}$ and Loss$_{SoftmaxLoss}$ in the formula. Where N represents the batch size, regression loss is calculated by the Euclidean distance, y_i represents the predicted aesthetic score, y_i represents the true aesthetic score, and f(w,x,b) represents the output of the upper layer of the category loss layer.

4 Experiments

In this section, we present the experimental results in the different dataset. First, we use the spatial perception aggregated deep convolutional neural network to use the training set on the IDEA for experimentation and the performance on the IDEA test set, the AVA test set, and the AADB test set. Evaluation. Next we trained and compared AADB datasets using spatially perceptually aggregated deep convolutional neural networks. Finally, we discuss the impact of fractional granularity partitioning on performance in multi-task learning (Fig. 6).

4.1 Implementation details

We fix the parameters of the layers before SPP layer of a pre-trained ResNet-50 model on the ImageNet [2] and fine tune all of the full connected layers on the training set of the AADB dataset and IDEA dataset. We use the Caffe framework [9] to train and test our models. The learning policy is set to step, The momentum of 0.9, a gamma of 0.5 and a weight decay of 0.0005. The training time is about four hours using Titan X Pascal GPU.

4.2 The performance of SAP-Net on AADB

To evaluate the aesthetic scores predicted by our model, we use the mean residual sum of squares error(MRSSE) and the ranking correlation measured by Spearman's between the estimated aesthetics scores and the ground-truth scores in the test set [18]. By comparison of

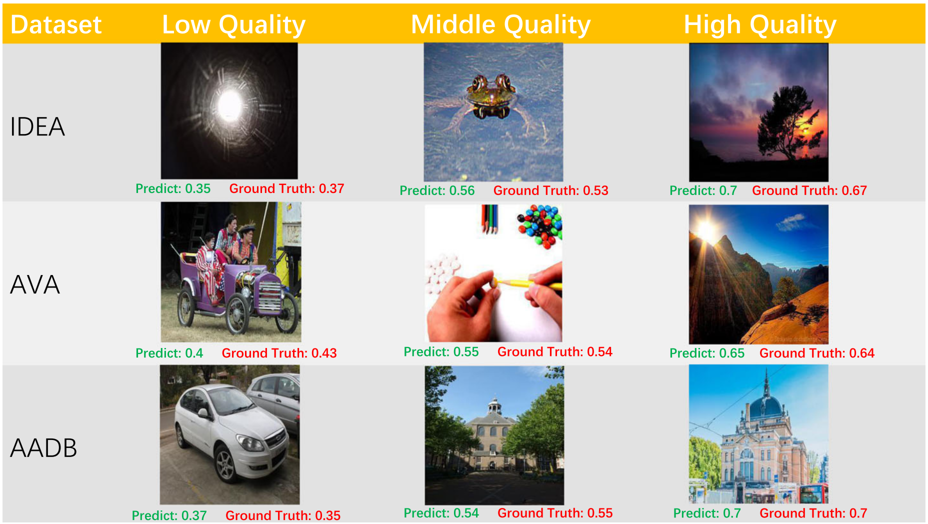

Dataset
Low Quality
Middle Quality
High Quality
IDEA
Predict: 0.35 Ground Truth: 0.37
Predict: 0.56 Ground Truth: 0.53
Predict: 0.7 Ground Truth: 0.67
AVA
Predict: 0.4 Ground Truth: 0.43
Predict: 0.55 Ground Truth: 0.54
Predict: 0.65 Ground Truth: 0.64
AADB
Predict: 0.37 Ground Truth: 0.35
Predict: 0.54 Ground Truth: 0.55
Predict: 0.7 Ground Truth: 0.7

Fig. 6 Some experimental results. Here we are on the IDEA data set for training, and test on the three data sets of validation set

Tables 1 and 2, it can be found that through the training of the AADB data set (the second-to-last row in the table), in the test of AVA's 20,000 test sets, the SAP-net can perform better performance. The performance of the SAP-net in MRSSE and the performance was significantly better than the previous methods. In addition, in the AADB verification set, Kong et al. [18] has the best performance of = 0.6782, and the SAP-net on the AADB validation set is equal to 0.6850, which is better than the previous method.

Obviously, it can be found that by using Squeeze-and-Excitation block module, the network can make the network more orderly. The control of overall accuracy is ensured by setting up auxiliary tasks. In addition, in terms of data sets, the training data on the IDEA dataset obviously have a stronger generalization ability, which is stronger than the AVA dataset on the AADB test set. Specifically, by modeling the dependence of each channel to improve the network representation, the network can adjust the features by channel so that the network can learn to selectively strengthen the features of the useful information and

Table 1 Performance comparison of aesthetic quality assessment on validation set of the AVA dataset using MRSSE

Methods	MRSSE
Always predicting 5 as aesthetic score	0.5700
BOV-SIFT+rbfSVR ([28] adapted in [15])	0.5515
BOV-SIFT+linSVR ([28] adapted in [15])	0.5401
GIST+rbfSVR ([28] adapted in [15])	0.5307
GIST+linSVR ([28] adapted in [15])	0.5222
Aest-CNN[15]	0.4501
DeepIA[1]	0.3727
SAP-Net (Training using AADB)	0.3146
SAP-Net (Training using IDEA)	0.2856

Table 2 Performance comparison of aesthetic quality assessment on validation set of the AVA dataset using

Methods	
Murray et al. [29]	–
SPP [4]	–
AlexNet FT Conf[18]	0.4807
DCNN [19]	–
RDCNN [19]	–
RDCNN semantic [20]	–
DMA [21]	–
DMA AlexNet FT [21]	–
Reg+Rank+Att+Cont[18]	0.5581
SAP-Net(Training using AADB)	0.5834
SAP-Net(Training using IDEA)	0.6159

suppress the useless characteristics through the global information. In this way, the network has stronger representation and learning ability. In addition, the network optimization space is controlled by the constraint conditions of the fractional section in multi task learning, thus the network makes the network effectively predicts each fraction effectively in training, thus maintaining a good MRSSE loss.

4.3 The performance of SAP-Net on IDEA

The AADB data set is better distributed than the AVA data set, and SAP-Net's learning ability is more fully developed in the IDEA data set. In the training of the IDEA training set (the last line in the Table 2), the performance of the SAP-Net in MRSSE and the performance was the best.

The training on IDEA data sets is better, which can be considered from the following two aspects. First, IDEA data sets make the distribution of network prediction more balanced, which is better for high and low score, while most methods of AADB and AVA data sets can only predict middle fraction well, and the scores of both ends are difficult to train because of small amount of data. Second, SAP-net has a stronger ability to express, by training more balanced data, it can give full play to the advantages of the network, so that the network has a stronger generalization ability (Table 3).

4.4 The effect of classification granularity on results

In order to explore the effect of classification granularity on the final regression results, we divided the scores into four categories in the training process of IDEA data sets. When

Table 3 Performance comparison of aesthetic quality assessment on validation set of the IDEA dataset using different classification granularity

The number of bins	
10	0.4962
5	0.5296
3	0.5722
2	0.4918

bin=10, the score is more discretized. When bin=2, the score is divided into two parts, which is equivalent to the binary classification problem. Through experiments, it can be found that at the time of bin=3, SAP-Net has the best performance in IDEA validation.

5 Conclusions

This paper proposes a new aesthetic uniform distribution data set (IDEA) and a new spatial perceptually aggregated deep convolutional neural network architecture. The low statistical property of IDEA datasets allows the neural network learning images to effectively learn deeper aesthetic features. Compared to other data sets, the uniform distribution of IDEA gives the network more aesthetic information. Through experiments, it is found that the results of training on IDEA have better performance on other data sets. Compared to traditional convolutions, SAP-net does not limit the input size of the network, reduces the loss of aesthetic information from resizing, and perceptually learns different channels after convolution extraction, effectively retaining more in-depth aesthetics feature. The network finally spatially aggregates the characteristics of these channels and optimizes the final results through multi-tasking objectives, so that the network has better performance in the experiment. There are many blind spots and difficulties in aesthetic evaluation. Aesthetic evaluation is an interdisciplinary subject of computer vision and many humanities sciences. There are more interesting discoveries waiting for people to explore. It is precisely these that keep us moving in this field.

Acknowledgements This work is partially supported by the National Natural Science Foundation of China (Grant Nos. 61772047, 61772513), the Science and Technology Project of the State Archives Administrator (Grant No. 2015-B-10), and the Fundamental Research Funds for the Central Universities (Grant No. 328201803, 328201801).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

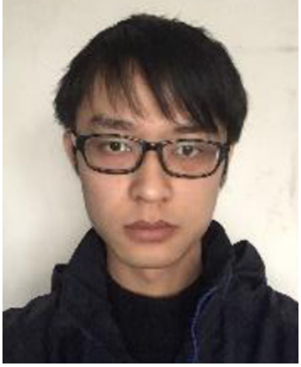
1. Bianco S, Celona L, Napoletano P et al (2016) Predicting image aesthetics with deep learning[C]. In: International conference on advanced concepts for intelligent vision systems. Springer, Cham, pp 117–125
2. Deng J, Dong W, Socher R et al (2009) Imagenet: a large-scale hierarchical image database[C]. In: CVPR 2009 IEEE conference on computer vision and pattern recognition, 2009. IEEE, pp 248–255
3. Dong Z, Tian X (2015) Multi-level photo quality assessment with multi-view features[J]. Neurocomputing 168:308–319
4. He K, Zhang X, Ren S et al (2014) Spatial pyramid pooling in deep convolutional networks for visual recognition[C]. In: European conference on computer vision. Springer, Cham, pp 346–361
5. He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
6. He L, Xu X, Lu H et al (2017) Unsupervised cross-modal retrieval through adversarial learning[C]. In: IEEE international conference on multimedia and expo. IEEE, pp 1153–1158
7. Hou L, Yu CP, Samaras D (2016) Squared earth mover's distance-based loss for training deep neural networks[J]. arXiv:1611.05916
8. Hu J, Shen L, Sun G (2017) Squeeze-and-excitation networks[J]. arXiv:1709.01507
9. Jia Y, Shelhamer E, Donahue J et al (2014) Caffe: convolutional architecture for fast feature embedding[C]. In: Proceedings of the 22nd ACM international conference on multimedia, pp 675–678. ACM

10. Jin B, Segovia MVO, Süsstrunk S (2016) Image aesthetic predictors based on weighted cnns[C]. In: 2016 IEEE international conference on image processing (ICIP). IEEE, pp 2291–2295
11. Jin X, Chi J, Peng S et al (2016) Deep image aesthetics classification using inception modules and fine-tuning connected layer[C]. In: 2016 8th international conference on wireless communications signal processing (WCSP). IEEE, pp 1–6
12. Jin X, Wu L, Song C et al (2017) Predicting aesthetic score distribution through cumulative Jensen-Shannon Divergence[C]. In: Proceedings of the 32th international conference of the America association for artificial intelligence (AAAI18), New Orleans, Louisiana, February 2-7, 2018
13. Kao Y, He R, Huang K (2017) Deep aesthetic quality assessment with semantic information[J]. IEEE Trans Image Process 26(3):1482–1495
14. Kao Y, Huang K, Maybank S (2016) Hierarchical aesthetic quality assessment using deep convolutional neural networks[J]. Signal Process Image Commun 47:500–510
15. Kao Y, Wang C, Huang K (2015) Visual aesthetic quality assessment with a regression model[C]. In: 2015 IEEE international conference on image processing (ICIP). IEEE, pp 1583–1587
16. Karayev S, Trentacoste M, Han H et al (2013) Recognizing image style[J]. arXiv:1311.3715
17. Ke Y, Tang X, Jing F (2006) The design of high-level features for photo quality assessment[C]. In: 2006 IEEE computer society conference on computer vision and pattern recognition, vol 1, pp 419–426. IEEE
18. Kong S, Shen X, Lin Z et al (2016) Photo aesthetics ranking network with attributes and content adaptation[C]. In: European conference on computer vision. Springer, Cham, pp 662–679
19. Lu X, Lin Z, Jin H et al (2014) Rapid: Rating pictorial aesthetics using deep learning[C]. In: Proceedings of the 22nd ACM international conference on multimedia. ACM, pp 457–466
20. Lu X, Lin Z, Jin H et al (2015) Rating image aesthetics using deep learning[J]. IEEE Trans Multimed 17(11):2021–2034
21. Lu X, Lin Z, Shen X et al (2015) Deep multi-patch aggregation network for image style, aesthetics, and quality estimation[C]. In: Proceedings of the IEEE international conference on computer vision, pp 990–998
22. Lu H, Li Y, Mu S et al (2017) Motor anomaly detection for unmanned aerial vehicles using reinforcement learning[J]. IEEE internet of things journal
23. Lu H, Li Y, Chen M et al (2017) Brain intelligence: go beyond artificial intelligence[J]. Mobile Networks and Applications, pp 1–8
24. Lu H, Li B, Zhu J et al (2017) Wound intensity correction and segmentation with convolutional neural networks[J]. Concurr Computat Pract Exper 29(6):e3927
25. Lu H, Li Y, Uemura T et al (2018) Low illumination underwater light field images reconstruction using deep convolutional neural networks[J]. Future Generation Computer Systems
26. Ma S, Liu J, Chen CW (2017) A-lamp: adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment[J]. arXiv:1704.00248
27. Mai L, Jin H, Liu F (2016) Composition-preserving deep photo aesthetics assessment[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 497–506
28. Marchesotti L, Perronnin F, Larlus D et al (2011) Assessing the aesthetic quality of photographs using generic image descriptors[C]. In: 2011 IEEE international conference on computer vision (ICCV). IEEE, pp 1784–1791
29. Murray N, Marchesotti L, Perronnin F (2012) A large-scale database for aesthetic visual analysis[C]. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR), pp 2408–2415. IEEE
30. Serikawa S, Lu H (2014) Underwater image dehazing using joint trilateral filter[J]. Comput Electr Eng 40(1):41–50
31. Wang W, Zhao M, Wang L et al (2016) A multi-scene deep learning model for image aesthetic evaluation[J]. Signal Process Image Commun 47:511–518
32. Wang Z, Liu D, Chang S et al (2017) Image aesthetics assessment using Deep Chatterjee's machine[C]. In: 2017 international joint conference on neural networks (IJCNN). IEEE, pp 941–948
33. Wu O, Hu W, Gao J (2011) Learning to predict the perceived visual quality of photos[C]. In: 2011 IEEE international conference on computer vision (ICCV). IEEE, pp 225–232
34. Xu X, He L, Lu H et al (2018) Deep adversarial metric learning for cross-modal retrieval[J]. World Wide Web-internet & Web Information Systems, pp 1–16

Xin Jin was born in Anhui province, China. He received the Ph.D. degree from Beihang University, China. Now, he is a lecture in Beijing Electronic Science and Technology Institute, China. His research is focused on visual computing and visual media security.

Le Wu was born in Shandong province, China. He received the Bachelor's degree from Shandong Normal University, China. Now, he is a post-graduate student in Beijing Electronic Science and Technology Institute, China. His research interest is visual media aesthetics.

Geng Zhao was born in Sichuan province, China. He received the Ph.D. degree from University of Science and Technology Beijing. He is a professor in Beijing Electronic Science and Technology Institute, China. His research interests include chaotic secure communications and computer information security.



Xinghui Zhou was born in Jiangxi province, China. He received the Bachelor's degree from Tianjin University, China. Now, he is a post-graduate student in Beijing Electronic Science and Technology Institute, China. His research interest is visual media aesthetics.

Xiaokun Zhang was born in Gansu province, China. He is now a professor and leader of the Department of Computer Science and Technology, Beijing Electronic Science and Technology Institute, China. His research interests include computer science and technology.

Xiaodong Li was born in Henan province, China. He received the Ph.D. degree from Northwestern Polytechnic University, China. Now, he is an associative professor in Beijing Electronic Science and Technology Institute, China. His research is focused on information security and visual media security.