





Predicting Aesthetic Radar Map Using a Hierarchical Multi-task Network

Xin Jin^{1,2} , Le Wu¹, Xinghui Zhou¹, Geng Zhao¹, Xiaokun Zhang¹,
Xiaodong Li¹, and Shiming Ge³ 

¹ Department of Cyber Security,
Beijing Electronic Science and Technology Institute, Beijing 100070, China
jinxin@besti.edu.cn

² CETC Big Data Research Institute Co., Ltd., Guiyang 550018, Guizhou, China

³ Institute of Information Engineering,
Chinese Academy of Sciences, Beijing 100093, China
geshiming@iie.ac.cn

Abstract. The aesthetic quality assessment of images is a challenging work in the field of computer vision because of its complex subjective semantic information. The recent research work can utilize the deep convolutional neural network to evaluate the overall score of the image. However, the focus in the field of aesthetic is often not limited to the total score of image, and multiple attribute of the aesthetic evaluation can obtain image richer aesthetic characteristics. The multi-attribute rating called Aesthetic Radar Map. In addition, traditional deep learning methods can only be predicted by classification or simple regression, and cannot output multi-dimensional information. In this paper, we propose a hierarchical multi-task dense network to make multiple regression of the properties of images. According to the total score, the scoring performance of each attribute is enhanced, and the output effect is better by optimizing the network structure. Through this method, the more sufficient aesthetic information of the image can be obtained, which is of certain guiding significance to the comprehensive evaluation of image aesthetics.

Keywords: Aesthetic evaluation · Neural network · Computer vision

1 Introduction

Recently, deep convolutional neural network technology has made great progress in the field of computer vision, especially object recognition and semantic recognition. However, the aesthetic quality of using computer to identify or evaluate images is far from practical. Image Aesthetic Quality Assessment (IAQA) is still a challenging task [1], the reasons are: large-scale data set of aesthetic is less in this field, aesthetic features are difficult for learning and generalization, evaluation of human subjectivity, etc. The aesthetic quality evaluation of images is a



Fig. 1. Aesthetic radar map and other assessment methods.

hot topic in the field of computer vision, computational aesthetics and computational photography.

In terms of the data set we use the PCCD aesthetic data set to train proposed by Chang et al. [22], which provided 7 kinds of aesthetic characteristics of the image, and we use these characteristics to compute the multiply scores. As shown in Fig. 1, according to the Aesthetic Radar Map we can get more complete and multi-angle evaluation aesthetic information. We will think it is a very good photo by scoring one number or classification, but it has some disadvantages in focus and exposure, which is very important for people’s aesthetic understanding, and the general one score regression or classification can not implement.

This paper presents a new hierarchical multi-task dense network architecture. Compared with the traditional learning method, this network can be strengthened from both global and attribute scoring, and finally get the total score of the image and the score of each attribute. In the feature extraction part of the convolution neural network, this paper use dense block structure [20] with different aesthetic characteristics in learning step, to reduce the phenomenon of vanishing-gradient and strengthens the use and transfer of feature information, and reduce the numbers of parameters to a certain extent. Behind the network part, we combine the study of the characteristics of global score and attribute score by fusion connection operation, to realize the global score effective utilization, and strengthens the attribute. Finally, through the combination of loss function, the network performs better. In the experimental part, this paper makes a comparison between the simple regression model and the non-hierarchical multi-task method, and proves that the proposed network and method have better performance. The main contributions of this paper are as follows:

- This is the first time to put forward the concept of the Aesthetic Radar Map and it fully show the aesthetic features with the Aesthetic Radar Map;
- Use the structure of the dense block in the aesthetic task to return the aesthetic score;
- For the first time, multi-task regression learning is applied to the aesthetic task, and a new feature fusion strategy is proposed to make the network selectively extract aesthetic features.

This paper predicts that the multi-attribute scoring of image aesthetic quality can be used for aesthetic image retrieval, photography technical guidance, video cover automatic generation and other applications. 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 have the eyes of beauty can we serve the human beings better.

2 Related Work

As mentioned in [2], the early work of image aesthetic quality evaluation mainly focuses on the manual design of various image aesthetic features and uses pattern recognition algorithm to make aesthetic quality prediction. Another research route tries to directly fit the quality of image aesthetics with some hand-designed universal image features. Recently, the study from big data depth image characteristics shows good performance [3–15], and the performance beyond the traditional manual design features. The training data for image aesthetic quality assessment usually comes from the online professional photography community, such as photo.net and dpchallenge.com. People can rate photos on these sites (1–7 or 1–10). The higher the score means the higher the aesthetic quality of the image [17].

Although aesthetic quality evaluation exists in a certain sense, it is still an inherent subjective visual task. The quality evaluation of image aesthetics is ambiguous [18], and there are different methods for quality evaluation of aesthetic images.

In the field of aesthetic classification, people usually use two value labels, such as good image and bad image, which are usually used to represent the quality of image aesthetics. In the field of aesthetic scoring, some regression network begins to get the score aesthetics of image, these models designed by convolution neural network to present image aesthetic quality of binary classification results or one-dimensional numerical evaluation [16, 23, 24]. Before the depth of neural network and mass aesthetic image quality evaluation dataset AVA [19] release, such as Wu et al. [17] training on small data sets, which is proposed based on support vector machine (SVM) prediction methods of the aesthetic image quality evaluation of distribution. Jin et al. [14] began to put forward an aesthetic histogram to better represent aesthetic quality, and Chang et al. [22] began to perform aesthetic image caption.

On aesthetic data set, Murray et al. [19] first puts forward the most massive data sets in aesthetics field, AVA, and gaussian distribution to fitting all the AVA data samples, the rest of the image evaluation scores can better be gamma distribution fitting [19]. Then, in view of the imbalance of AVA samples, Kong et al. [12] proposed the AADB data set to make the aesthetic data set more balanced and better proper in the normal distribution. Chang et al. [22] proposed the PCCD data set, which is a relatively comprehensive small-scale data set.

3 Hierarchical Multi-task Network

3.1 Aesthetics Radar

For aesthetic image evaluation, the evaluation of a score is often incomplete. Through the evaluation of the pictures through several aesthetic indicators, a more comprehensive and a richer evaluation can be obtained. Usually such evaluation is also more meticulous.

The data set we use is called PCCD. It is based on the evaluation of the basic score, in the meantime, it considered the influence of Subject of Photo, Composition & Perspective, Use of Camera, Exposure & Speed, Depth of Field, Color & Lighting, Focus on the evaluation of the picture is also considered, and finally it is plotted in the form of a radar chart.

The composition of the picture evaluation will be updated from low dimension to high dimension, and some of the features with clear features can also be well represented by radar charts (Fig. 2).

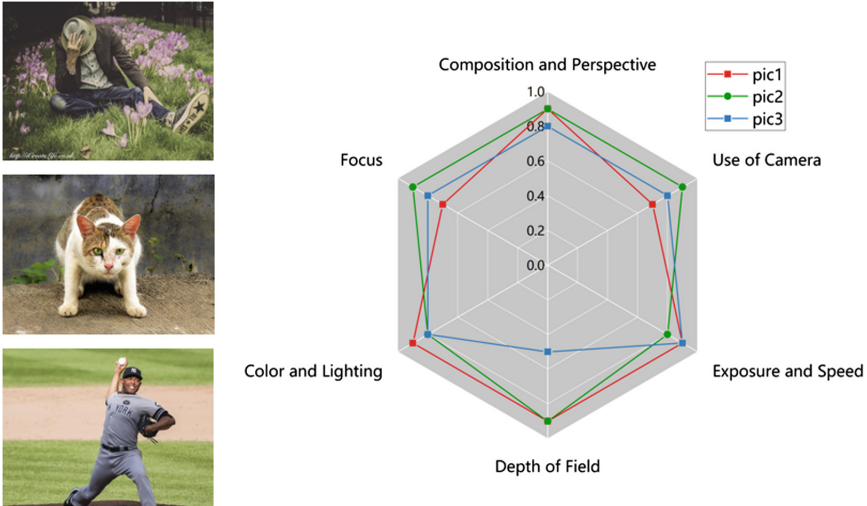


Fig. 2. Samples in the Photo Critique Captioning Dataset (PCCD)

The PCCD (Photo Critique Captioning Dataset) data set is a model for verifying the problems arising from the proposed aesthetic image evaluation, provided by Chang et al. [22]. The dataset is based on the professional photo review website¹ and provides experienced photographers' comments on the photos. On the website, photos were displayed and some professional reviews were provided in the following seven areas: general impressions, composition and perspective, color and lighting, photo theme, depth of field, focus and camera usage, exposure and speed.

¹ <https://gurushots.com/>.

3.2 Dense Module

The dense module neural network was proposed in CVPR2017 [20]. Its algorithm is based on ResNet [21], but its network structure is completely new. Dense module can effectively reduce the number of features in a neural network while achieving better results. In each Dense Model, the input for each layer comes from the output of all previous layers. At the same time, each layer can relate to the input data and the loss, which can alleviate over-fitting and the problem of gradient disappearing when the network is too deep (Fig. 3).

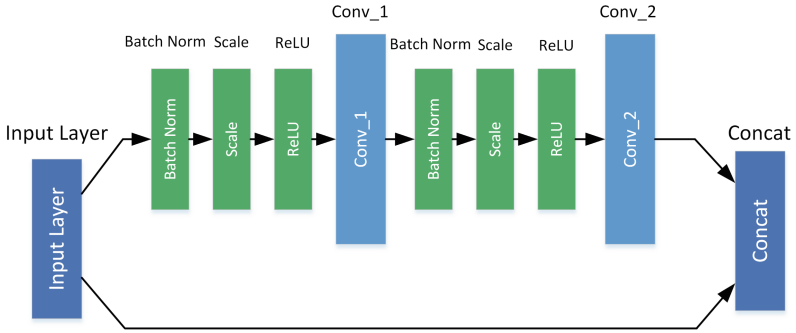


Fig. 3. Dense module

In ResNet, the relationship between two adjacent layers can be expressed by the following formula:

$$X_l = H_l(X_{l-1}) + X_{l-1} \quad (1)$$

where l denotes the layer, X_l denotes the output of layer l , and H_l denotes a nonlinear transform. So for ResNet, the output of layer l is the output of layer $l - 1$ plus the nonlinear transformation of the output of layer $l - 1$.

By changing the way information is transmitted between layers, dense module proposes a new connection method. Any one of them needs to relate to its subsequent layer. Its mathematical expression is as follows:

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]) \quad (2)$$

where $[X_0, X_1, \dots, X_{l-1}]$ refers to the concatenation of the feature-maps produced in layers $0, \dots, l - 1$ (Fig. 4).

There H_l as a composite function of three consecutive operations: batch normalization (BN), a rectified linear unit (ReLU) and a convolution (Conv). Due to the dense connectivity of the network, we refer to this network architecture as a dense convolutional network (DenseNet).

Dense module produces k output maps for each layer, but there are more inputs. In a specific application, a 1×1 convolution is added as a bottleneck

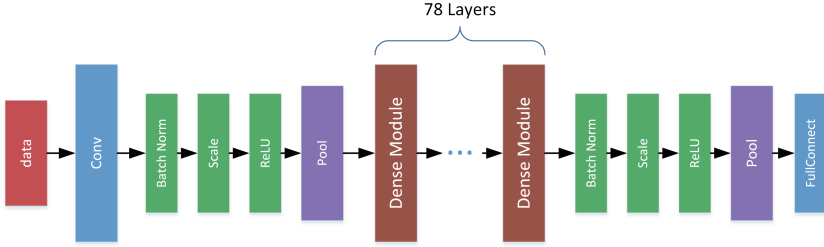


Fig. 4. The structure of feature extract network

before each 3×3 convolution to reduce the number of input feature maps, thereby increasing the computational efficiency. We have found that this design is particularly effective for dense module, and this method has been the bottleneck in the network.

3.3 Hierarchical Multi-task

Multi-task learning (MTL) is a common algorithm widely used in machine learning and deep learning. Due to the diversity of its results, MTL can achieve multi-angle evaluation of picture aesthetics through parameter sharing. The results of picture evaluation under different angles are relatively independent, but the model training process is the same. The Hierarchical MTL structure used in the experiment like Fig. 5.

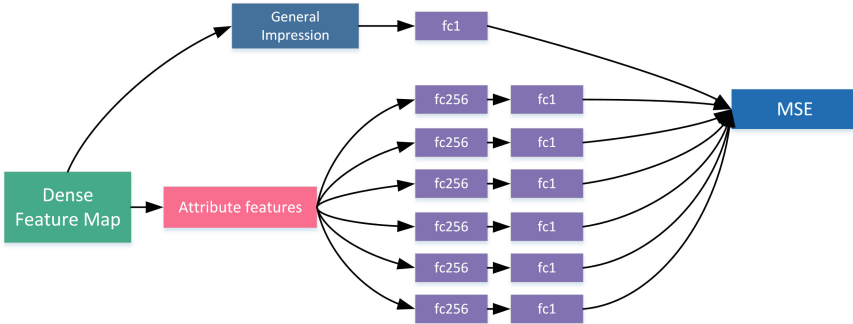


Fig. 5. The multi-task part of HMDnet (hierarchical multi-task dense network)

The dense module output at the last full-connection level is divided into seven parts, general impression and another six aesthetic attributes. Next, we split six aesthetic properties on the output by full-connection operation and perform the same operation to create the general impression. For the final result,

the calculation of the mean-square error (MSE) is performed and returned as a model loss parameter to the previous network.

Hierarchical multi-task is a joint learning method. It learns multiple attributes of a picture, solves multiple problems at the same time, and performs regression prediction on multiple problems. A typical Multi-task, for example, in the business area, the personalized problem, from analysing multiple hobbies of a person to get a more comprehensive evaluation plan.

Hierarchical multi-task image processing methods have two advantages over traditional statistical methods:

- The radar image can display multi-angled and multi-leveled image information. In this experiment, pictures often have different levels of picture attributes and can be vividly represented by Multi-task;
- Multi-task evaluation pictures are often more specific and detailed. Multi-task analysis pictures can show the advantages and disadvantages of the picture in all aspects.

4 Experiment

4.1 Implementation Details

We fix the parameters of the layers before the first full connected layer of a pre-trained densenet model on the ImageNet [2] and fine-tune the all full connected layers on the training set of the PCCD dataset. We use the Keras framework² to train and test our models. The learning policy is set to step. Stochastic gradient descent is used to train our model with a mini-batch size of 16 images, a momentum of 0.9, a learning rate of 0.001 and a weight decay of 1e−6. The max number of iterations is 160. The training time is about 40 min using Titan X Pascal GPU.

4.2 Predict Result

For the data output by our model, dimension reduction is performed through the full connect layer, and regression calculations are performed on the known scores to obtain the predicted values of six aesthetic attributes of a picture and a total score estimate. The size of the Test data set is 500 pictures.

The experimental prediction results and test dataset data fitting results are better. Among them, the Color and Lighting attribute and the Composition and Perspective attribute have better results, and the other four attributes have larger deviations. The overall result is accurate. Some predict demo shown in Fig. 6.

² <https://github.com/keras-team/keras/>.

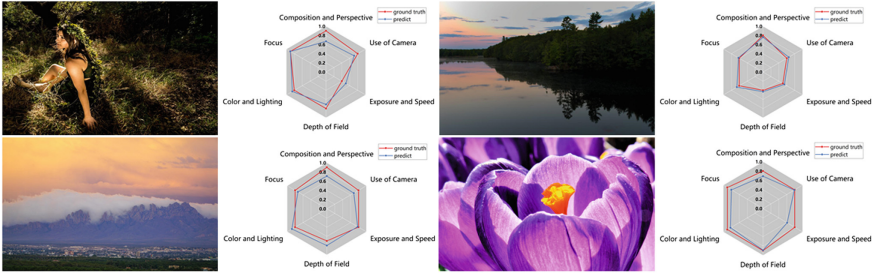


Fig. 6. Predicted results of test data set photos and ground truth.

4.3 Compare with Other Methods

To verify the effectiveness of our experimental results, we compared the algorithm (HMDNet) with other algorithms. The regression method uses densenet to make a simple regression to the score, without adding multi-attribute and multilayer full-connection structure, multi-task method uses multi-attribute combination method but does not use the total score. For the same data set, we get a better fit for the model predictions and the real data. Compared with other methods, we can prove that our method has more advantages in multi-task picture aesthetic reviews.

Table 1. The predictions’ MSE of HMDNet and other methods.

Methods	GI	SP	CP	UES	DF	CL	FO
Regression	0.086801	0.13978	0.109241	0.111274	0.204511	0.122637	0.223453
Multi-task	0.079941	0.14742	0.094143	0.127399	0.150707	0.094961	0.173752
HMDNet	0.079646	0.12789	0.076158	0.109694	0.128662	0.088098	0.142878

As Shown in Table 1, the GI means General Impression, it’s a general evaluate of a picture. The SP which in the Table 1 means Subject of Photo, the CP means Composition & Perspective, the UES means Use of Camera, Exposure & Speed, the DF means Depth of Field, the CL means Color & Lighting, the FO means Focus. Our methods can get best performance in overall score and all attribute scores.

5 Conclusions

This paper puts forward a new Hierarchical Multitasking convolution neural network architecture. We present a new aesthetic task and goal of Aesthetic Radar Map, and predict it through the multi-task regression network. Compared with the traditional regression network, this paper makes full use of the

global aesthetic rating to make the overall score and attribute rating interact with each other, thus realizing the accurate prediction of multi-attribute tasks. Experiments show that this method makes the prediction closer to the real label. As an interdisciplinary subject of computer vision, photography and iconography, aesthetic evaluation has more interesting discoveries waiting for people to explore, and many blind areas await our in-depth discovery.

Acknowledgments. We thank all the reviewers and ACs. This work is partially supported by the National Natural Science Foundation of China (grant numbers 61772047, 61772513, 61402021), the open funding project of CETC Big Data Research Institute Co.,Ltd., (grant number W-2018022), the Science and Technology Project of the State Archives Administrator (grant number 2015-B-10), the Open Research Fund of Beijing Key Laboratory of Big Data Technology for Food Safety (grant number BTBD-2018KF-07), Beijing Technology and Business University, and the Fundamental Research Funds for the Central Universities (grant numbers. 328201803, 328201801).

References

1. Mai, L., Jin, H., Liu, F.: Composition-preserving deep photo aesthetics assessment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 497–506 (2016)
2. Deng, J., Dong, W., Socher, R., et al.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
3. Karayev, S., Trentacoste, M., Han, H., et al.: Recognizing image style. arXiv preprint [arXiv:1311.3715](https://arxiv.org/abs/1311.3715) (2013)
4. Lu, X., Lin, Z., Jin, H., et al.: RAPID: rating pictorial aesthetics using deep learning. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 457–466. ACM (2014)
5. Kao, Y., Wang, C., Huang, K.: Visual aesthetic quality assessment with a regression model. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 1583–1587. IEEE (2015)
6. Lu, X., Lin, Z., Shen, X., et al.: Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 990–998 (2015)
7. Lu, X., Lin, Z., Jin, H.: Rating image aesthetics using deep learning. *IEEE Trans. Multimed.* **17**(11), 2021–2034 (2015)
8. Dong, Z., Tian, X.: Multi-level photo quality assessment with multi-view features. *Neurocomputing* **168**, 308–319 (2015)
9. Kao, Y., Huang, K., Maybank, S.: Hierarchical aesthetic quality assessment using deep convolutional neural networks. *Sig. Process. Image Commun.* **47**, 500–510 (2016)
10. Wang, W., Zhao, M., Wang, L.: A multi-scene deep learning model for image aesthetic evaluation. *Sig. Process. Image Commun.* **47**, 511–518 (2016)
11. Ma, S., Liu, J., Chen, C.W.: A-Lamp: adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. CoRR abs/1704.00248. URL: <http://arxiv.org/abs/1704.00248> (2017)

12. Kong, S., Shen, X., Lin, Z., Mech, R., Fowlkes, C.: Photo aesthetics ranking network with attributes and content adaptation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 662–679. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_40
13. Jin, X., Chi, J., Peng, S., et al.: Deep image aesthetics classification using inception modules and fine-tuning connected layer. In: 2016 8th International Conference on Wireless Communications Signal Processing (WCSP), pp. 1–6. IEEE (2016)
14. Jin, X., Wu, L., Song, C., et al.: Predicting aesthetic score distribution through cumulative jensen-shannon divergence. In: Proceedings of the 32th International Conference of the America Association for Artificial Intelligence (AAAI 2018), New Orleans, Louisiana, 2–7 February 2018 (2017)
15. Kao, Y., He, R., Huang, K.: Deep aesthetic quality assessment with semantic information. IEEE Trans. Image Process. **26**(3), 1482–1495 (2017)
16. Wang, Z., Liu, D., Chang, S., et al.: Image aesthetics assessment using Deep Chatterjee’s machine. In: 2017 International Conference on Neural Networks (IJCNN), pp. 941–948. IEEE (2017)
17. Wu, O., Hu, W., Gao, J.: Learning to predict the perceived visual quality of photos. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 225–232. IEEE (2011)
18. Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 419–426. IEEE (2006)
19. Murray, N., Marchesotti, L., Perronnin, F.: AVA: a large-scale database for aesthetic visual analysis. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2408–2415. IEEE (2012)
20. Iandola, F., Moskewicz, M., Karayev, S., et al.: DenseNet: implementing efficient convnet descriptor pyramids. arXiv preprint [arXiv:1404.1869](https://arxiv.org/abs/1404.1869) (2014)
21. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
22. Chang, K.Y., Lu, K.H., Chen, C.S.: Aesthetic critiques generation for photos. In: 2017 IEEE International Conference on Computer Vision (ICCV) pp. 3534–3543. IEEE (2017)
23. Jin, B., Segovia, M.V.O., Süssstrunk, S.: Image aesthetic predictors based on weighted CNNs. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 2291–2295. IEEE (2016)
24. Hou, L., Yu, C.P., Samarasinghe, D.: Squared earth mover’s distance-based loss for training deep neural networks. arXiv preprint [arXiv:1611.05916](https://arxiv.org/abs/1611.05916) (2016)