https://doi.org/10.2352/ISSN.2470-1173.2020.8.IMAWM-188

Semi-supervised Multi-task Network For Image Aesthetic Assessment*

Xiaoyu Xiang[a], Yang Cheng[a], Jianhang Chen[a], Qian Lin[b], Jan Allebach[a]
[a] School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA
[b] HP Labs. Palo Alto, CA 94304, USA

Abstract

Image aesthetic assessment has always been regarded as a challenging task because of the variability of subjective preference. Besides, the assessment of a photo is also related to its style, semantic content, etc. Conventionally, the estimations of aesthetic score and style for an image are treated as separate problems. In this paper, we explore the inter-relatedness between the aesthetics and image style, and design a neural network that can jointly categorize image by styles and give an aesthetic score distribution.

To this end, we propose a multi-task network (MTNet) with an aesthetic column serving as a score predictor and a style column serving as a style classifier. The angular-softmax loss is applied in training primary style classifiers to maximize the margin among classes in single-label training data; the semi-supervised method is applied to improve the network's generalization ability iteratively. We combine the regression loss and classification loss in training aesthetic score. Experiments on the AVA dataset show the superiority of our network in both image attributes classification and aesthetic ranking tasks.

Introduction

Image aesthetic assessment is essential to smart photo album applications, such as the criteria for image retrieval systems and cover photo recommendation. Besides, it can serve as a guide for image enhancement. For a long time, image aesthetic assessment has been regarded as a branch of image quality assessment, which includes the estimation of several image quality attributes, such as sharpness, noise, artifacts, etc. While the goal of image quality assessment algorithms is to be consistent with the quality assessment of a human viewer [23], image aesthetics cannot be fully represented by image quality alone. Sometimes the human assessment of aesthetics even contradicts the assessment of quality for the same image. Compared with other computer vision problems, image aesthetic assessment is even more challenging because of its subjective nature [26], which means we can hardly find a universal rule to judge a given image.

In recent years, convolutional neural networks have been widely applied in image aesthetics prediction [2, 7, 9, 10, 14, 21, 24, 27] thanks to the publishing of large-scale datasets, such as AVA [22], AADB [13], and IDEA [11]. These datasets are composed of image data and meta-data, such as crowd-sourcing scores, semantic categories, and style or attributes. For human viewers, although there is no strict rule that directly corresponds to aesthetic scores, they are perceptually influenced by following factors in a given image: the theme, or semantic elements within the image, and the proper style and composition, such as the rule of thirds, shallow DOF (depth of field), etc. In other words, the aesthetics of an image is assessed through a complex interplay between themes and styles. Moreover, images with obvious styles are usually linked to high aesthetic appeal.

Inspired by the intrinsic connection between style and aesthetic assessment, this work entangles these two problems together by a multi-task network (MTNet), that can be trained to encode given images into feature vectors and predict aesthetic scores, as well as the style category, where the feature vectors are shared by an aesthetic column and a style column. For efficient training, we first pre-train the encoder on the AVA training set with A-softmax loss [16], and iteratively fine-tune it by a self-training approach. The acquired encoder serves as a backbone for MTNet. Overall, the proposed MTNet is shown to improve the performance over the previous state-of-the-art architectures.

Related Works

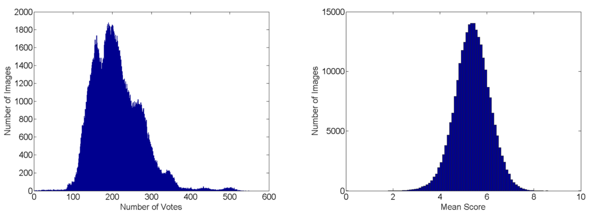
Aesthetic score prediction

Deng *et al* [4] summarizes early works using hand-crafted aesthetic-rule based features to design an assessment method [3, 15, 25]. More recent studies show that using a deep feature representation method boosted by large-scale datasets performed much better than traditional hand-crafted features [10]. The training data are often collected from online photography communities, where people rate each photo by integer scores from 1 to 10. By weighted averaging all scores, it should be possible to get the final score for an image. Higher scores represent a positive assessment from most of the people. Based on the score, one can compare two images in terms of their aesthetic aspects, and classify an image into a high or low aesthetic category.

Some researchers [27] choose to represent image aesthetic quality with the score, either in numerical encoding or binary encoding. However, other researchers propose that due to the subjective nature of human annotators, the score distribution can be really diverse. Especially, the mean score is easily influenced by low and high extremes from minor annotators. Besides, previous papers [9, 10, 11] also notice that the distribution of mean score or the number of the votes for each image is far from evenly distributed as shown in Figure 1. In order to solve this problem, more recent works propose to use the score distribution in image aesthetic assessment. The progress on this path is mainly to use more and more appropriate loss functions to make the network's output approach the target distribution. Jin *et al* [9] applies the weighted Chi-square distance loss to predict the average score and standard

*Research supported by HP Inc., Palo Alto, CA.

deviation from the rankings distribution. Wang *et al* uses asymmetrical Kullback-Leibler (KL) divergence as the loss function and processes the AVA dataset's score distribution into a Gaussian distribution to train their DBN network. Hou *et al* [8] and Talebi *et al* [24] uses squared EMD (earth mover's distance) loss [8] to train a CNN to predict the score histogram. Murray *et al* [21] employs the Huber loss, which is more robust to outliers, in their APM network. Cui *et al* [2] applies the traditional label distribution learning method to predict an aesthetic score histogram. Jin *et al* [10] compares previously proposed loss functions and chooses the cumulative Jensen-Shannon divergence.



(a) (b)

Figure 1. *The score and vote number distribution of the AVA dataset: (a) The distribution of the number of votes per image; (b) The distribution of mean score per image.*

Image style classification

Early works [5] use hand-crafted features to describe photo styles. Lu *et al* [17] utilizes style attributes of the image to help improve the aesthetic classification accuracy, where the style-SCNN is pretrained on the AVA dataset and its output is concatenated to predict the binary categorization result. Inspired by the multi-column and concatenation methods, Kong *et al* [13] incorporates joint learning of photographic attributes and image content to predict image aesthetic scores.

Embedding of the MTNet

Network Architecture

The architecture of MTNet is shown in Figure 2. It can be generally divided into two parts: an image encoder, and a multi-column network to realize specific tasks, including an aesthetic column that outputs the score distribution for the image's aesthetics, and an attribute column that classifies the attributes of the given image. The image encoder roughly follows the same structure as SphereFace [16], which is a modified ResNet. Input images are encoded into a 512-dimension feature vector for further processing in multi-column networks.

The attribute column is composed of a fully connected layer that converts the 512×1 feature vector into a 14-class prediction result. Since we apply angular-softmax (A-softmax) to train the classifier, the predicted score for each class ranges from -1 to 1.

Angular-softmax

A-softmax loss is a novel loss function modified from softmax loss, first proposed by [16]. Traditional softmax loss is commonly used for classification tasks. For an input feature vector x_i and its corresponding label y_i, the softmax function can be formulated as follows:

$$L_{softmax} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{z_{y_i}}}{\sum_{j=1}^{C} e^{z_j}} \tag{1}$$

where N is the number of training samples, C stands for the number of classes. z_j is the activation of the $j-$th neuron in a fully connected layer with corresponding weight vector W_j and bias b_j. Each neuron is supposed to predict the score z_j for target class j. Softmax loss is based on the Euclidean margin of different classes. If we fix the bias $b_j = 0$ and express x in the polar coordinate system, the predicted score z_j can be written as:

$$z_j = W_j^T x = ||W_j|| \cdot ||x|| \cos\theta_j \tag{2}$$

where θ_j denotes the angle between W_j and the input x. Previous papers [6, 16] use a simplified example of binary classification to illustrate the idea of A-softmax. According to Equation 2, three components can influence the predicted score: an embedding feature vector x, and the learned weights W_1 and W_2, which represent the class centers. For this binary classification problem, the decision boundary is:

$$||W_1|| \cos\theta_1 = ||W_2|| \cos\theta_2 \tag{3}$$

For a single label problem, a sample is classified into the class j that has a higher score z_j. As shown in Equation 3, the decision boundary depends on both the norm and angle of the classification center. Even if the angles between a feature vector and two class centers are the same, it tends to be categorized to the class with a larger L2 norm. Besides, a feature vector with a larger L2 norm would output a higher prediction score.

In this paper, we want to eliminate the influence of norms and classify over angularly discriminative features (The reason to disentangle the L2 norm is illustrated in the Appendix). We normalize $||W_j||$ and $||x||$ to 1 as follows:

$$W_j = \frac{W_j^*}{||W_j^*||}, \quad x = \frac{x^*}{||x^*||} \tag{4}$$

where W_j^* and x denote the original weight and feature vectors. After normalization, only the angles between the feature vector and the two weight vectors are relevant to the decision boundary:

$$\cos\theta_1 = \cos\theta_2 \tag{5}$$

That is, the sample tends to be classified into the class j that has larger $\cos\theta_j$, which corresponds to a smaller angle. We can regard the normalization as mapping the weight vector and feature vector to a unit hypersphere. As analyzed above, $||W_j|| = 1, \forall j$ and $||x|| = 1$. The modified A-softmax is:

$$L_{ang} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{m\cos\theta_{y_i}}}{\sum_{j=1}^{C} e^{m\cos\theta_j}} \tag{6}$$

where m is the scale factor that controls the size of the margin. From softmax to A-softmax loss, our network learns image features with an angular margin. A-softmax has already demonstrated its superiority in many tasks including face recognition and verification [16], person re-identification [6], and other general classification tasks [28].

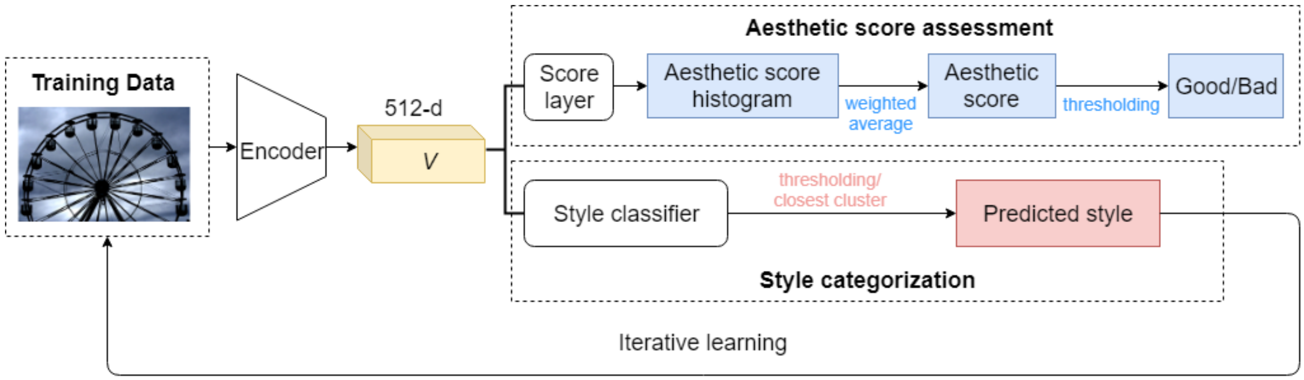


Figure 2. *Structure of the MTNet.*

Semi-supervised Training

In most computer vision research, data augmentation including random cropping, re-centering, RGB normalization, and noise-injection is used as an effective tool to improve the network's robustness. However, cropping and re-centering would change the composition of an image. This would lead to totally different aesthetic rankings for human perception. And RGB normalization would make the acquired network less sensitive to brightness, white balance, etc, which happen to be important factors for image aesthetic assessment. Thus, methods of data augmentation should be carefully chosen for image aesthetic related research projects. In this paper, we horizontally flipped 50% of training images to ensure the least influence on the data's inner consistency. All images are resized to 112×96 to satisfy the requirements of the network input.

In the AVA attributes subset, all training images have a single label, while the number of labels of test images can vary from zero to many. The difference between the training and testing data requires our network to have the ability to generalize, while maintaining discrimination between multiple attribute classes. A semi-supervised strategy is used to achieve this goal. First, we train the primary classifier with A-softmax loss until it actually learns discriminative attribute features (reaching 99.9823% accuracy on the training set in our experiment). Then we apply this pre-trained classifier on training data again to predict the attribute scores for each image. With the reference of ground truth, we add the attribute classes for which the prediction scores are higher than a certain threshold into the training labels. After scanning the whole dataset, we get a new training set with multi-label attributes. Then, we re-train the classifier with the acquired labels and re-apply it to the training set again to propose new attribute labels to training images. We repeat this process until the classifier stops changing. The iterative learning algorithm is summarized in Figure 3.

After the iterative learning process, we are supposed to have a robust classifier that can distinguish image styles. This classifier can be divided into an image encoder and an attribute column. Assuming that the encoded feature vectors have enough information reflecting image attributes, it is reasonable to expect that information about aesthetic scores is also included in the latent features. The following section introduces how to train an aesthetic column based on this assumption.

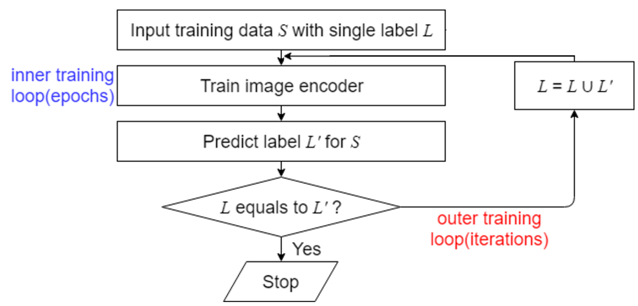


Figure 3. *The diagram of the iterative learning process.*

Combining the Regression and the Classification

The target output of the aesthetic column is the score histogram as mentioned in the earlier introduction of the AVA dataset. On the one hand, the aesthetic score prediction task is firstly a regression problem that aims to approach the target score as closely as possible. The mean square error (MSE) loss is widely adopted for solving such regression tasks. On the other hand, the predicted scores serve as a reference for binary classification, that categorizes the input image into either "good" or "bad" categories. Cross entropy loss, softmax loss, etc. are used to solve this classification problem.

In this paper, we combine the regression problem with the classification problem to get better results. By adding extra layers to the aesthetic column, our network is capable of outputting a score histogram and binary classification scores. For these two types of outputs, we apply the MSE loss and the Cross-Entropy loss, respectively, to the two tasks. The advantage of combining the two losses is to increase distinctions among training data. As shown in Figure 1, the score histogram of the AVA dataset is not evenly distributed. Previous research [11] has declared that he unbalanced training set makes the training of neural networks easy to over-fit. The adding of classification loss teaches the network to be more discriminative on the images near the threshold.

Experiments

In training the attribute classifier, we choose $m = 4$. Stochastic gradient descent is used as the optimizer, with a momentum of 0.9 and a weight decay of 0.0005. We fix the parameters of the

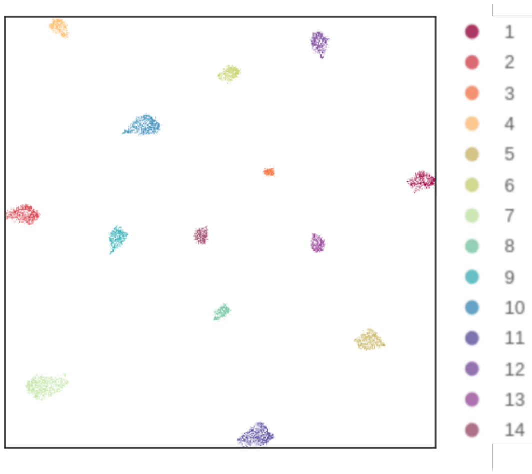


Figure 4. The attributes of photos in the AVA datasets embedded into two dimensions by the UMAP.

image encoder when training the aesthetic column, and fine-tune the remaining layers with AVA's training set. During the fine-tuning, the optimizer's weight decay is changed to 0.005 while others remain the same.

We perform two types of evaluation: the classification accuracy using the benchmark-setting on AVA style experiments [22], and divergence between the predicted score histogram and the ground truth in AVA aesthetic image lists [22] using the same metrics as described in [10].

Style Classification

To prove that our image encoder actually learns the features representing image attributes and scores, we visualize the encoded features by dimension reduction using UMAP (Uniform Manifold Approximation and Projection) [19, 20], as shown in Figure 4. For simplicity, we plot the AVA attribute training set where each image only corresponds to one label. We can see that the attributes of different labels form separate clusters in the hyperspace, which indicates our network is able to learn a discriminative embedding of the latent features of the AVA images.

We compare our model's performance on the test set of AVA, where each image may correspond to 0 to n labels. Following the approach taken in previous papers, we report AP (Average Precision), which computes the average precision value for recall values over 0 to 1, and mAP (mean Average Precision), which is the average AP for all categories.

Table 1 shows that our method achieves the second-best results for style classification. Here, we list our results acquired from three different methods: 1) primary method, which is the primary classifier that is trained on the original single-label dataset; 2) iterative learning model, which is acquired based on the primary classifier through the iterative learning process as mentioned previously; 3) nearest cluster method. It uses the model trained from iterative learning. But instead of taking the classification output from the last fully connected layer, it takes the 512-dimension feature vector to compute the cosine distance to the class's average centers of each style category. The classes with a distance that is lower than the threshold are taken as prediction outputs. In our experiment, the distance thresholding method outperforms the direct output of the fully-connected layer by 0.05%.

Aesthetic Score Histogram Prediction

Table 2 shows the comparison of score histogram prediction results on AVA dataset. The divergence between the predicted histogram and the ground truth is measured by several metrics: PED (the Euclidean distance between two probability distribution functions); PCE (cross-entropy between two probability distribution functions); PJS (the symmetrical version of the Jensen-Shannon divergence between two probability distribution functions); PCS (Chi-square distance between two probability distribution functions) [9]; PKL (the symmetrical version of Kullback-Leibler divergence between two probability distribution function) [26]; CED (Euclidean distance between two cumulative distribution functions) [8, 27]; and CJS (the symmetrical version of the Jensen-Shannon divergence between two cumulative distribution functions) [10]. With the exception of PCS, our result outperforms the previous best ones by a large margin.

Conclusion

In this paper, we present a multi-task deep convolutional neural network for image aesthetic quality assessment and style categorization. Rather than solving these problems with two separate models, in our MTNet the aesthetic column and the style column share the same image encoder trained by a semi-supervised approach, which provides a new insight for solving the multi-label classification problem with single-label data. Besides, we introduce A-softmax loss in the image aesthetics area for the first time, and demonstrate its effectiveness on style classification in an angular hyperspace. Our experiment results prove that the performance of aesthetic assessment can be leveraged by style classification. The evaluation results on the AVA dataset show that our approach outperforms earlier-reported methods for the six out of the seven metrics evaluated.

Appendix

Interpretation of L2 Norm

To understand what the L2 norm of feature vectors encoded by our network represents, we select the top three and bottom three pictures of each style according to their feature vectors' L2 norm. The selected pictures are shown in Figure 5.

The feature vector is a distilled representation of the original image. In our work, the network basically serves as a discriminator for good images with recognizable styles. Each dimension of the vector represents a detectable feature related to the network's purpose, which implies that a higher norm corresponds to stronger detectable features in the image. The pictures in Figure 5 support this: images with higher norms usually have richer content and more distinguishable features. In order to get rid of the bias caused by these factors of the L2 norm for style classification, we chose A-softmax loss along with cosine distance as a comparison metric.

References

[1] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006.

Method	AP (Average Precision)	mAP (mean Average Precision)
Murray *et al* [22]	n/a	53.85%
Karayev *et al*[12]	n/a	58.1%
Lu *et al*[17]	56.93%	56.81%
DMA-Net [18]	**69.78%**	**64.07%**
Kong *et al*[13]	59.11%	58.73%
Ours (primary)	60.75%	58.89%
Ours (iterative learning)	61.19%	59.47%
Ours (class distance)	62.03%	60.28%

Table 1: Image style classification results on the AVA testet. Bold font indicates the best result, and underline font indicates the second-best result.

Method	PED	PCE	PJS	PCS	PKL	CED	CJS
PED	0.197	2.830	0.059	0.105	0.728	0.323	0.068
RS-PED	0.189	2.733	0.055	0.094	0.657	0.324	0.067
PCE	0.167	2.773	0.041	0.075	0.442	0.279	0.049
RS-PCE	0.169	2.771	0.046	0.071	0.438	0.279	0.047
PJS	0.185	2.828	0.051	0.093	0.527	0.326	0.053
RS-PJS	0.183	2.776	0.049	0.091	0.523	0.327	0.049
PCS[9]	0.182	2.807	0.045	0.082	0.450	0.287	0.045
RS-PCS	0.175	2.783	0.045	0.079	0.423	0.277	0.044
PKL[26]	0.163	2.779	0.039	0.073	0.389	0.270	0.044
RS-PKL	0.164	2.778	0.037	0.071	0.386	0.268	0.043
MMD[1]	0.201	2.831	0.064	0.112	0.710	0.339	0.068
RS-MMD	0.196	2.824	0.063	0.097	0.710	0.322	0.054
Huber[21]	0.184	2.775	0.044	0.078	0.409	0.279	0.053
RS-Huber	0.183	2.774	0.045	0.074	0.402	0.271	0.048
CED[8][27]	0.182	2.799	0.047	0.085	0.502	0.294	0.049
RS-CED	0.180	2.792	0.048	0.082	0.502	0.283	0.047
CJS[10]	0.163	2.779	0.039	0.072	0.382	0.266	0.041
RS-CJS[10]	0.158	2.760	0.037	**0.068**	0.381	0.260	0.040
LDL Method[2]	0.303	-	-	-	-	-	-
Ours	**0.060**	**2.218**	**0.009**	0.081	**0.018**	**0.110**	**0.012**

Table2: Comparison of score histogram prediction errors on the AVA dataset. Smaller numbers indicate better results. The results are obtained by training with different loss functions as mentioned in [10]. Bold font indicates the best result.

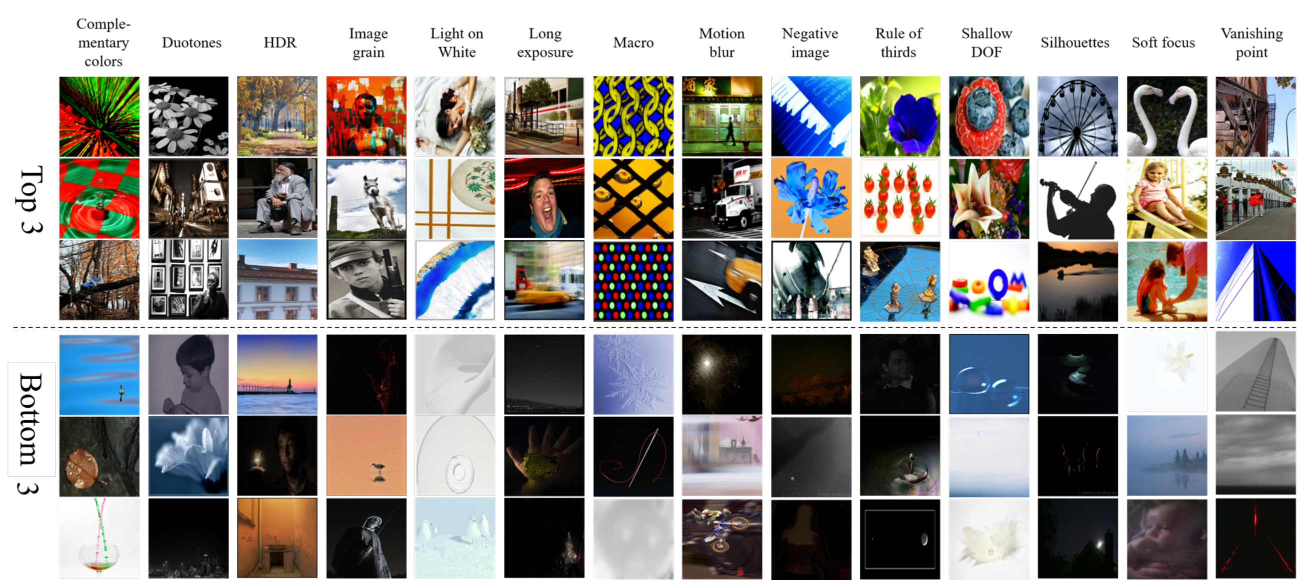


Figure 5. *Images selected by feature vectors' norm for each style*

[2] Chaoran Cui, Huidi Fang, Xiang Deng, Xiushan Nie, Hongshuai Dai, and Yilong Yin. Distribution-oriented aesthetics assessment for image search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1013–1016. ACM, 2017.

[3] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *European Conference on Computer Vision*, pages 288–301. Springer, 2006.

[4] Yubin Deng, Chen Change Loy, and Xiaoou Tang. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine*, 34(4):80–106, 2017.

[5] Sagnik Dhar, Vicente Ordonez, and Tamara L Berg. High level describable attributes for predicting aesthetics and interestingness. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1657–1664. IEEE, 2011.

[6] Xing Fan, Wei Jiang, Hao Luo, and Mengjuan Fei. Spherereid: Deep hypersphere manifold embedding for person re-identification. *Journal of Visual Communication and Image Representation*, 60:51–58, 2019.

[7] Xin Fu, Jia Yan, and Cien Fan. Image aesthetics assessment using composite features from off-the-shelf deep models. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3528–3532. IEEE, 2018.

[8] Le Hou, Chen-Ping Yu, and Dimitris Samaras. Squared earth mover's distance-based loss for training deep neural networks. *arXiv preprint arXiv:1611.05916*, 2016.

[9] Bin Jin, Maria V Ortiz Segovia, and Sabine Süsstrunk. Image aesthetic predictors based on weighted CNNs. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2291–2295. IEEE, 2016.

[10] Xin Jin, Le Wu, Xiaodong Li, Siyu Chen, Siwei Peng, Jingying Chi, Shiming Ge, Chenggen Song, and Geng Zhao. Predicting aesthetic score distribution through cumulative Jensen-Shannon divergence. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[11] Xin Jin, Le Wu, Geng Zhao, Xinghui Zhou, Xiaokun Zhang, and Xiaodong Li. Idea: A new dataset for image aesthetic scoring. *Multimedia Tools and Applications*, pages 1–15, 2018.

[12] Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. Recognizing image style. *arXiv preprint arXiv:1311.3715*, 2013.

[13] Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. Photo aesthetics ranking network with attributes and content adaptation. In *European Conference on Computer Vision*, pages 662–679. Springer, 2016.

[14] Kwan-Yee Lin and Guanxiang Wang. Self-supervised deep multiple choice learning network for blind image quality assessment. In *The British Machine Vision Conference*, page 70, 2018.

[15] Ligang Liu, Renjie Chen, Lior Wolf, and Daniel Cohen-Or. Optimizing photo composition. In *Computer Graphics Forum*, volume 29, pages 469–478. Wiley Online Library, 2010.

[16] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 212–220, 2017.

[17] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James Z Wang. Rapid: Rating pictorial aesthetics using deep learning. In *Proceedings of the 22nd ACM International Conference on Multimedia*, pages 457–466. ACM, 2014.

[18] Xin Lu, Zhe Lin, Xiaohui Shen, Radomir Mech, and James Z Wang. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 990–998, 2015.

[19] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*, February 2018.

[20] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861, 2018.

[21] Naila Murray and Albert Gordo. A deep architecture for unified aesthetic prediction. *arXiv preprint arXiv:1708.04890*, 2017.

[22] Naila Murray, Luca Marchesotti, and Florent Perronnin. AVA: A large-scale database for aesthetic visual analysis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2408–2415. IEEE, 2012.

[23] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[24] Hossein Talebi and Peyman Milanfar. NIMA: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8):3998–4011, 2018.

[25] Hanghang Tong, Mingjing Li, Hong-Jiang Zhang, Jingrui He, and Changshui Zhang. Classification of digital photos taken by photographers or home users. In *Pacific-Rim Conference on Multimedia*, pages 198–205. Springer, 2004.

[26] Zhangyang Wang, Ding Liu, Shiyu Chang, Florin Dolcos, Diane Beck, and Thomas Huang. Image aesthetics assessment using deep Chatterjee's machine. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 941–948. IEEE, 2017.

[27] Ou Wu, Weiming Hu, and Jun Gao. Learning to predict the perceived visual quality of photos. In *2011 International Conference on Computer Vision*, pages 225–232. IEEE, 2011.

[28] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. Robust classification with convolutional prototype learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3474–3482. IEEE, 2018.

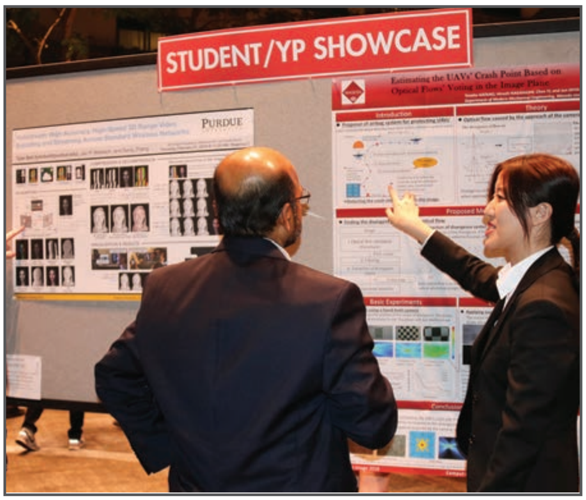
JOIN US AT THE NEXT EI!

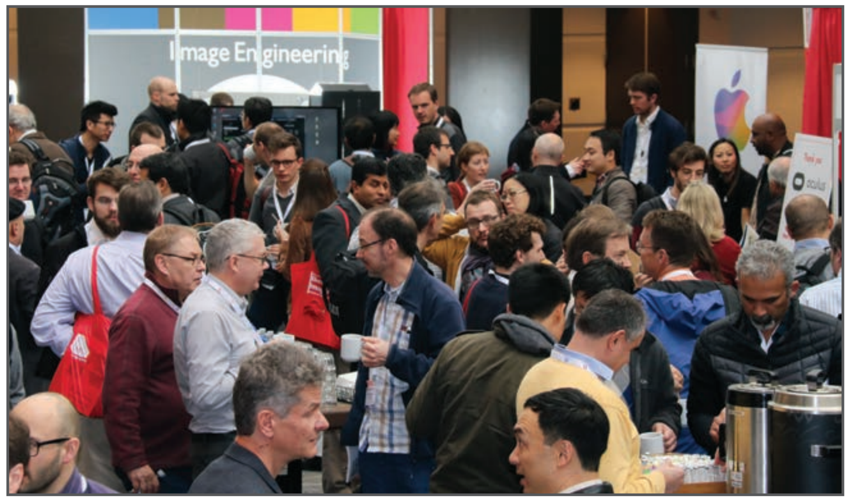
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org