ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Bin Jin, Maria V. Ortiz-Segovia, Sabine Süsstrunk

IMAGE AESTHETIC PREDICTORS BASED ON WEIGHTED CNNS

Do you like the crops?

Aesthetics and CNNs

Automatically assessing image aesthetics is useful for many applications, such as image retrieval, image enhancement and images album management, as shown in the automatic cropping example above. State-of-the-art aesthetics prediction algorithms [1-4] train Convolutional Neural Networks (CNNs) on a recently-published large-scale aesthetics dataset, AVA [5].

Dataset

AVA dataset: ~250K images with ~200 aesthetic score ratings per image, ranging from 1(low) to 10(high). We split the AVA dataset into one training set and two different test sets: the ***RS-test***, 3000 Random Sampled images, and the ***ED-test***, 3000 Evenly Distributed images among three categories: : low quality (aesthetic score <4), medium quality (4≦aesthetic score ≦7), and high quality (aesthetic score >7).

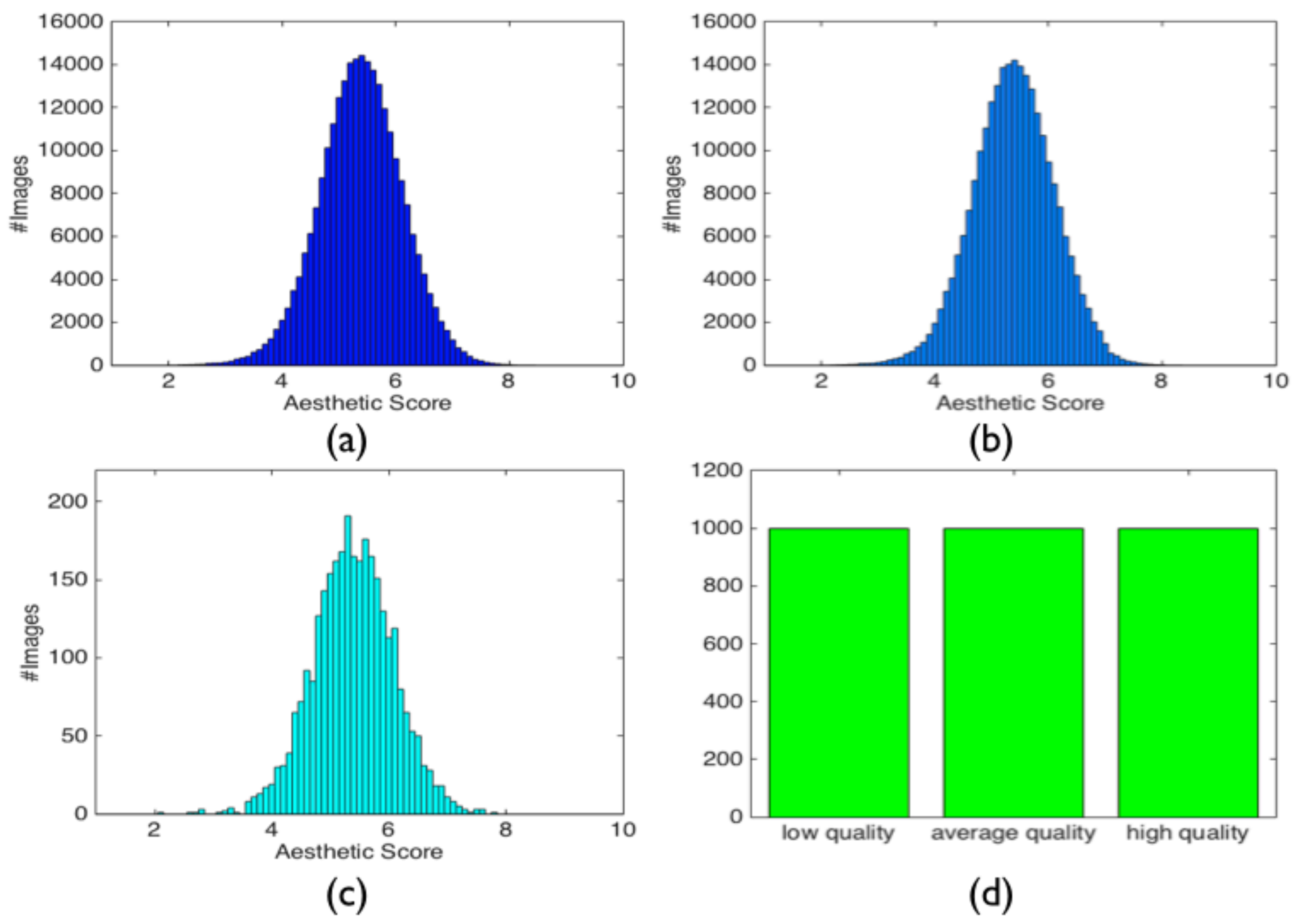


Fig. 2. The distribution of the average aesthetic scores for (a) the whole AVA dataset (b) the training set, (c) the *RS-test*, (d) the *ED-test*.

Our Regression Model

The regression model is trained to predict **the average aesthetic score**. By using the Weighted Mean Squared Error (WMSE) loss function, we manage to reduce the bias introduced by the unbalanced training set.

$$b_i' = \frac{b_i}{\sum_{i=1}^{B} b_i}; \qquad w_i = \frac{1}{b_i'}$$

$$WMSE = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \cdot (y_i - \hat{y}_i)^2$$

Here y_i is the predicted aesthetic score and $\hat{y}_i$ is the groundtruth aesthetic score. N is the number of images in the training set. w_i is the weight of each sample. B is the total number of bins in the aesthetic scores histogram and b_i is the occurrences number of the ith bin.

Our Histogram Prediction Model

The histogram prediction model aims at predicting **the normalized histogram of user ratings** for an input image. Two values can be derived from the histogram output, the average aesthetic score, which is the predicted aesthetic level, and the standard deviation of user ratings, which represents the difficulty of aesthetics assessment. The loss function for training is Weighted χ^2 Mean Error (WMCE):

$$WMCE = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \cdot \chi^2(\mathbf{h_i}, \hat{\mathbf{h}}_\mathbf{i})$$

Here w_i is the sample weight for image i, $\mathbf{h_i}$ is the output histogram from the network and $\hat{\mathbf{h}}_\mathbf{i}$ is the groundtruth normalized histogram, χ^2 represents the chi-square distance.

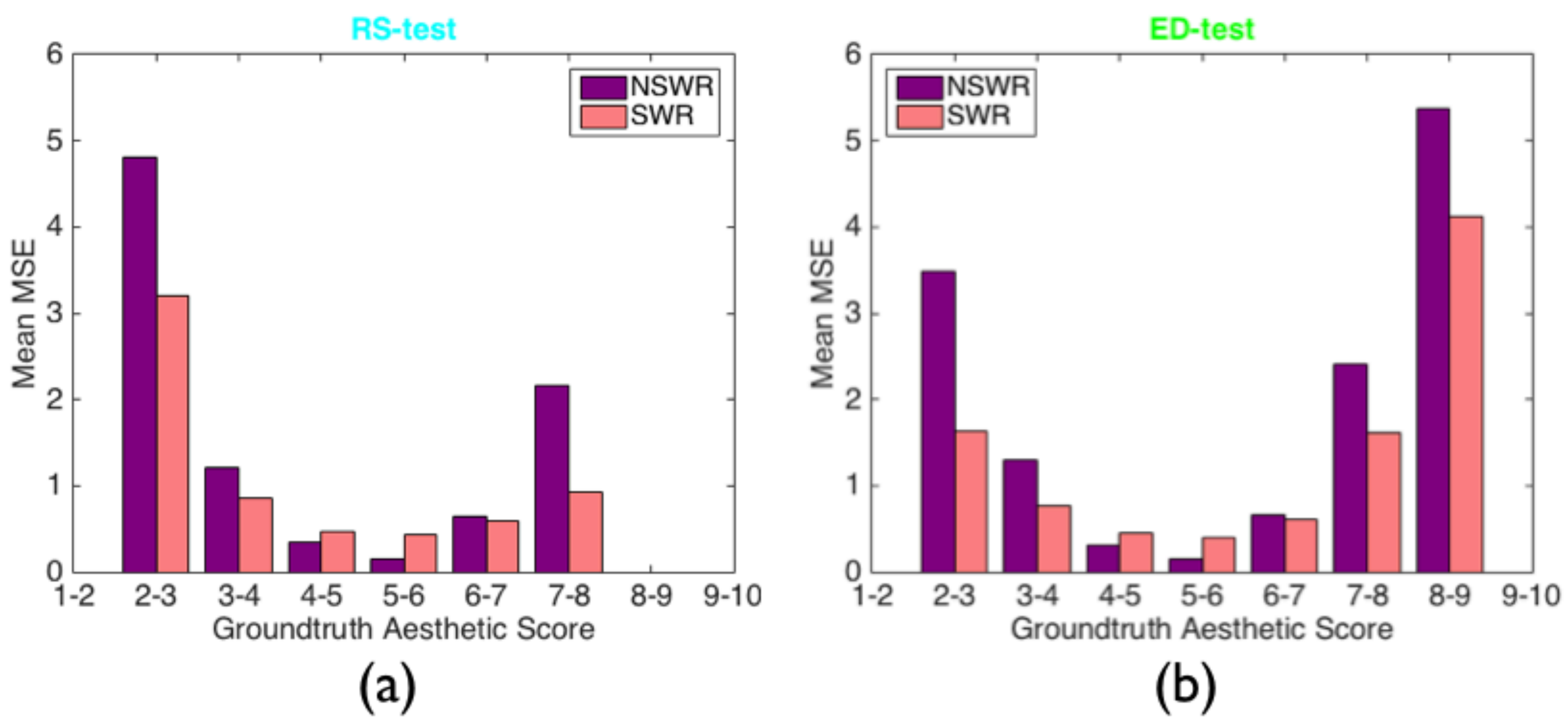
Results

Two regression CNN models with the same architecture are trained: a *Regression* model with ***No Sample Weights*** (***NSWR***) and a *Regression* model with ***Sample Weights*** (***SWR***). The Mean Square Error (MSE) for different models are shown in the Table 1.

Table. 1. MSE of different models.

	RS-test	*ED-test*
GIST linear-SVR	0.5222	NA
GIST rbf-SVT	0.5307	NA
BoVW SIFT linear-SVR	0.5401	NA
BoVW SIFT rbf-SVR	0.5513	NA
Kao et al.[1]	0.4510	NA
No Sample Weights (*NSWR*)	**0.3373**	1.3951
Sample Weights (*SWR*)	0.4847	**0.9754**

We further show the mean MSE for different aesthetic score ranges in Fig. 3. Using sample weights clearly contributes to reducing the MSE for images with aesthetic scores larger than 6 or smaller than 4.



For the histogram prediction model, the average aesthetic score and the standard deviation (std) are derived from the predicted histogram. We use MSE to evaluate the derived average aesthetic score and Root Mean Square Error Ratio (RMSER) to evaluate the std. *SWH* achieves comparable performance as the *SWR* for predicting the aesthetic scores on the ***ED-test***, while producing less than 20% RMSER. Hence, the difficulty of aesthetics assessment for an image is also reliably estimated.

Table. 2. Performance of the histogram prediction model.

	MSE	RMSER
RS-test	0.6358	26.75%
ED-test	1.0109	19.57%

Conclusion

We propose to use sample weights while training CNNs for aesthetics prediction. Our regression models and histogram prediction model demonstrate the effectiveness of the sample weights for reducing the bias in the training set.

References

[1] Yueying Kao, Chong Wang and Kaiqi Huang, "Visual aesthetic quality assessment with a regression model," in *Image Processing (ICIP), 2015 IEEE International Conference on*. 2015, pp. 1583–1587.
[2] Xin Lu, Zhe Lin, Hailin Jin, Xin Yang, 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*. 2014, pp. 457–466.
[3] Luca Marchesotti, Naila Murray and Florent Perronnin, "Discovering beautiful attributes for aesthetic image analysis," in *International Journal of Computer Vision*. 2014, vol. 113, no. 3, pp. 246–266.
[4] Florian Simond, Nikolaos Arvanitopoulos Darginis and Sabine Süsstrunk , "Image aesthetics depends on context," in *Image Processing (ICIP), 2015 IEEE International Conference on*. 2015, pp. 3788–3792.
[5] Naila Murray, Luca Marchesotti and Florent Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. 2012, pp. 2408– 2415.

This work is sponsored by océ Canon