Predicting Subjectivity in Image Aesthetics Assessment

Chen Kang, Giuseppe Valenzise, Frédéric Dufaux

Predicting Subjectivity in Image Aesthetics Assessment

Chen Kang, Giuseppe Valenzise, Frédéric Dufaux
Laboratoire des Signaux et Systèmes (L2S), CNRS-CentraleSupélec-Université ParisSud, Université Paris-Saclay
{chen.kang, giuseppe.valenzise, frederic.dufaux}@l2s.centralesupelec.fr

***Abstract*—Conventional image aesthetic quality prediction aims at predicting the average score of a picture or its aesthetic class (good/bad quality). However, aesthetic prediction is intrinsically subjective, and images with similar mean aesthetic scores/class might display very different levels of consensus by human raters. Recent work has dealt with aesthetic subjectivity by predicting the distribution of human scores. However, predicting the distribution is not directly interpretable in terms of subjectivity, and might be sub-optimal compared to directly estimating subjectivity descriptors computed from ground-truth scores. In this paper, we propose several measures of subjectivity, ranging from simple statistical measures such as the standard deviation of the scores, to newly proposed descriptors inspired by information theory. We evaluate the prediction performance of these measures when they are computed from predicted score distributions or when they are directly learned from ground-truth data. We find that the latter strategy provides in general better results, though there is still a large space for improvement in aesthetic subjectivity prediction.**

***Index Terms*—Aesthetic quality, subjectivity, distribution prediction**

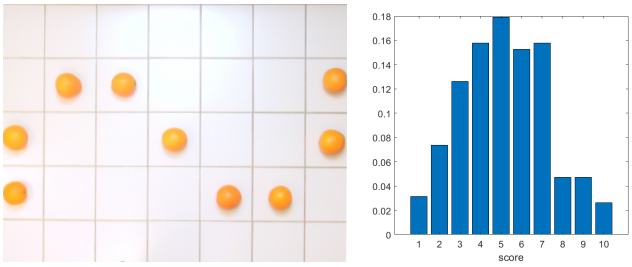
I. Introduction

The goal of image aesthetic quality assessment is to determine how beautiful an image looks to a human observer. The automatic prediction of image aesthetic quality has received an increasing attention in the past few years in the multimedia community. This is due, on one hand, to the potential impact that aesthetic quality prediction has on applications such as image enhancement, recommendation or retrieval [1]. On the other hand, the availability of large-scale datasets with human annotations [2], [3] has enabled the use of modern machine learning tools, such as deep learning, to predict aesthetic scores for images displaying a wide variety of contents and characteristics.

Most of existing aesthetic quality prediction approaches assume that aesthetic quality can be represented by a single value, e.g., the mean aesthetic score or the aesthetic class (good/bad). However, this assumption does not take into account the intrinsic *subjectivity* of aesthetic assessment, which may be influenced by personal background, interests, mood, etc. Indeed, experimental psychology studies show that, while beauty is conveyed by objective visual clues, the resulting aesthetic appraisal is subjective and depends on how the visual clues are processed by higher-level cognitive areas in the brain [4].

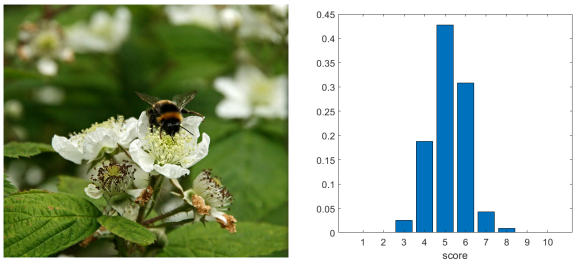
This work is founded by China Scholarship Council.

978-1-7281-1817-8/19/$31.00 ©2019 IEEE



Mean	SD	MAD	MED	DUD
5.174	2.104	1.697	0.278	3.570

(a)



Mean	SD	MAD	MED	DUD
5.180	0.902	0.657	0.650	4.439

(b)

Fig. 1. Example of aesthetic subjectivity for two images of the AVA dataset. The two images, displayed in the top-left panels, have similar mean score but different distribution of aesthetic judgments given by human raters, shown in the histograms on the top-right panels. The tables report several measures that compactly describe subjectivity based on the score distribution, which are described in Section III.

As a result, summarizing aesthetic quality with a single value is not in general sufficient to capture the subjectivity of aesthetic perception, which we define in this paper as the *degree of consensus* about the aesthetic value of a picture when the latter is judged by a panel of human raters. The top two rows of Figure 1 illustrate this with an example: two images from the AVA dataset have a similar average aesthetic score, but a different degree of subjectivity. In the image in Figure 1(a), it is evident that humans tend to agree more on the aesthetic quality of the image, while the judgments are more dispersed for the image in Figure 1(b). Intuitively, being able to predict aesthetic subjectivity can provide valuable information in order to determine to which extent aesthetic predictions can

be trusted. This in turn could be beneficial in applications such as enhancement or retrieval, in order to obtain more reliable and accurate results.

Recent work has tackled aesthetic subjectivity by predicting the *distribution* of subjective scores of an image [5]–[8]. Specifically, these methods leverage the availability of ground-truth aesthetic score distributions obtained by a large number of human annotations, offered by large-scale datasets such as AVA [2], and employ different loss functions to measure the distance between probability distributions. However, aesthetic subjectivity is described only implicitly by the score distribution. Instead, we are interested at quantifying and predicting this subjectivity *explicitly* through a scalar value that summarizes the score distribution by describing the raters' consensus, and which could be used by automatic image analysis algorithms.

Therefore, in this paper we analyze several measures of subjectivity based on the distribution of the scores, including simple statistical descriptors such as standard deviation, as well as new proposed features inspired by information theory. We evaluate the prediction accuracy of the proposed subjectivity measures using state-of-the-art aesthetic score distribution prediction, and compare these results with directly learning the subjectivity scores. Our experiments show that directly learning subjectivity measures leads, in general, to better performance than first predicting the score distribution and then computing subjectivity based on it. Despite the improvement obtained by our approach using several performance indicators, our findings show that predicting aesthetic subjectivity is a much more difficult task than predicting the average aesthetic score of a picture.

The rest of the paper is organized as follows. In Section II we review related work on image aesthetics prediction, and we present several subjectivity descriptors in Section III. We evaluate different prediction schemes for these measures in Section IV. Finally, Section V concludes the paper.

II. Related Work

General aesthetics prediction deals with predicting image aesthetics for any kind of image content, in contrast to task-specific aesthetics where the class or object of the picture is known, e.g., images of faces [9]. In this work we focus on the general image aesthetics problem. Traditional methods for predicting image aesthetics have been employing hand-crafted features to describe well-known photographic rules and perceptual attributes, such as clarity, depth of field, colorfulness, dynamic range, etc. [10]–[12]. These methods have the advantage to provide an interpretable explanation of image aesthetics, but fail in capturing accurately complex aesthetic phenomena. As a result, these approaches tend to perform poorly when tested in real-world conditions with a wide content variety.

With the advent of deep learning and the availability of large-scale datasets with thousands or hundreds of thousands pictures [2], [3], the accuracy of aesthetics prediction has been constantly improving. In this context, the problem of aesthetic assessment has been mainly formulated as predicting the average score or the aesthetic class of an image [1], [3], [13], [14].

On the other hand, the problem of subjectivity in image aesthetic quality assessment has been rarely studied. Park et al. [15] consider personal taste in addition to general aesthetic score, by adapting a model to match specific user preferences obtained from user interactions. Differently from that work, we do not target personalized aesthetic prediction, but rather aim at assessing the level of consensus of a panel of humans about the general aesthetic value of a picture.

Recently, a few studies have considered aesthetic subjectivity by predicting the distribution of human scores, rather than simply the mean score. Bin Jin et al. [5] predict the distribution of aesthetic scores based on a weighted loss which accounts for the non-uniform distribution of the scores in the AVA dataset, using a modified VGG-16 [16] network. Their loss function uses the chi-square distance to evaluate the distribution prediction. Later, Murray et al. [6] employ the Huber loss and spatial pyramid pooling [17] to predict distributions. In the NIMA system [7], the loss function consists of the squared Earth Mover's Distance (EMD), which is shown to lead to better mean score prediction performance from the estimated distributions. Jin et al. [8] propose to use the cumulative Jensen-Shannon Divergence (CJS-CNN) as loss function. They also present an extended version of this loss using a function of the kurtosis of ground-truth score distribution to weigh CJS (RS-CJS). Kurtosis is used as a proxy to aesthetic "reliability", and used to penalize more those images whose distribution is considered unreliable. In this paper, instead of predicting the score distributions, we propose (for the first time to our knowledge) to define explicitly subjectivity measures and directly predict them.

The majority of the above-mentioned works employ the benchmark AVA dataset [2] to learn distributions or average scores. It contains over 250,000 images from photography amateurs' websites, which is much larger comparing to previous datasets like CUHK-PQ [18], [19]. The images are collected by approximately 1400 challenges from viewers who voted integer scores in the range $[1, 10]$. Compared to other datasets like AADB [3] which only has around 5 voters for each image, the number of votes in AVA ranges between 78 and 549, and the average is around 210, thus enabling a more reliable estimation of score distributions. In this work, we also employ the AVA dataset to train and evaluate subjectivity prediction.

III. Prediction of Subjectivity

We consider a dataset of N images $\{I_n\}$, $n = 1 \dots N$, where each image has been voted by M_n human raters on a discrete scale with k levels, $s = \{s_1, \dots, s_k\}$. We model the M_n aesthetic scores x_n for each image I_n as a realization of a categorical random variable with distribution $p_n(x_n)$, which we approximate with the normalized sample histogram $\mathbf{p}_n(x_n)$. Given $\mathbf{p}_n(x_n)$, we define μ_n and m_n as the mean and median of x_n, respectively.

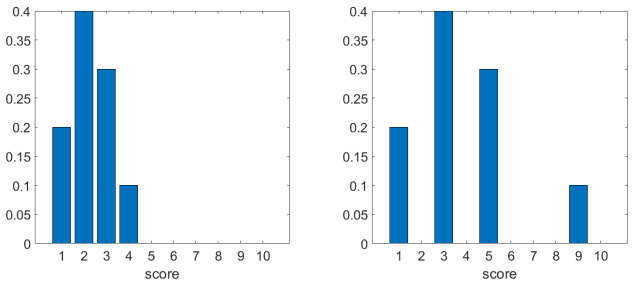


Fig. 2. The two score distributions have the same entropy, but the one on the right has a higher degree of subjectivity.

In order to describe the level of consensus of human raters about the aesthetic quality of a given image, we propose using the following measures:

- **Standard Deviation** (SD) of the score distribution, which describes the dispersion of the scores around the average score, that is:

$$SD_n = \sum_{i=1}^{k} p_n(i) \cdot (x_n(i) - \mu_n)^2. \tag{1}$$

  A higher value of SD indicates a lower consensus around the average score, and thus higher subjectivity.
- **Mean Absolute Deviation around the median** (MAD), defined as the sample average deviation of the scores around the median score, that is:

$$MAD_n = \frac{1}{M_n} \sum_{i=1}^{M_n} |x_n(i) - m_n|. \tag{2}$$

  As for SD, higher values of MAD imply higher subjectivity.
- **Distance to Uniform Distribution** (DUD). The entropy of a distribution characterizes the degree of uncertainty of the associated random variable, and could be in principle used to quantify subjectivity. However, entropy does not take into account the ordinal nature of aesthetic scores, as illustrated in Figure 2. Instead of measuring entropy, we consider the distance of the score distribution $\mathbf{p}_n(x_n)$ from the distribution having the maximum entropy over s, which is the uniform distribution. We quantify this distance using the 2-Wasserstein metric[1] $d_W(\mathbf{p}_n, \mathbf{u}_s)$, that is:

$$DUD_n = d_W(\mathbf{p}_n, \mathbf{u}_s) = \left[\sum_{i=1}^{k} (\mathbf{P}_n(i) - \mathbf{U}_s(i))^2 \right]^{1/2}, \tag{3}$$

  where $\mathbf{u}_s$ is the discrete uniform distribution defined over the categories s, and $\mathbf{P}_n$ and $\mathbf{U}_s$ are the cumulative distribution functions of $\mathbf{p}_n$ and $\mathbf{u}_s$, respectively.

[1]Note that the 2-Wasserstein metric is sometimes confused with the Earth Mover Distance, e.g., in [7]. However, for the sake of precision, the Earth Mover Distance corresponds to the 1-Wasserstein metric.

  A lower value of DUD implies that the score distribution is more similar to the uniform distribution, and thus the degree of subjectivity is higher.
- **Distance from the Maximum Entropy Distribution** (MED). Since the uniform distribution has always a mean value equal to the midpoint of the score scale, the DUD measure tends to penalize more skewed distributions having mean values close to the extremes of the quality scale. To overcome this bias, we compare the score distribution with the maximum entropy distribution over the quality scale having the *same mean*. More specifically, we look for a discrete distribution $\mathbf{q}_s$ which solves the following optimization problem:

$$\begin{aligned} \underset{\mathbf{q}}{\text{maximize}} \quad & H(\mathbf{q}) \\ \text{subject to} \quad & \mu[\mathbf{q}] = \mu_n, \end{aligned}$$

  where H denotes discrete entropy and $\mu[\mathbf{q}]$ is the mean of $\mathbf{q}$. It can be shown [20] that the solution of this problem is

$$\mathbf{q}_s(s_i) = \frac{e^{\lambda s_i}}{\sum_{i=1}^{k} e^{\lambda s_i}}, \tag{4}$$

  where λ is numerically found so that $\sum_i s_i \mathbf{q}_s(s_i) = \mu_n$. Then MED for image n is defined as:

$$MED_n = d_W(\mathbf{p}_n, \mathbf{q}_s) = \left[\sum_{i=1}^{k} (\mathbf{P}_n(i) - \mathbf{Q}_s(i))^2 \right]^{1/2}, \tag{5}$$

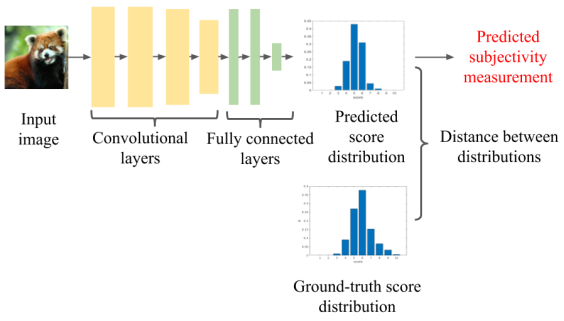
  where $\mathbf{Q}_s$ is the cumulative distribution of $\mathbf{q}_s$. As for the DUD measure, the lower MED is, the higher is the subjectivity of an image.

The table in Figure 1 shows an example of these measures computed for the two images in the top panel. We can observe that all of them capture correctly the degree of consensus of the score distributions. In the following, we will study how accurately each of these measures can be predicted, either directly or by means of predicted score distributions.

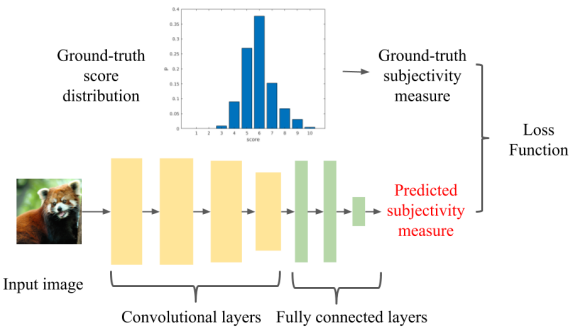
A. Subjectivity Prediction Framework

In order to predict the subjectivity measures proposed above, we consider two options: i) we predict the score distribution *indirectly* using an existing score prediction method as mentioned in Section II; or ii) we compute subjectivity measures on ground-truth scores, and learn to predict them *directly*.

Indirect subjectivity prediction: The underlying motivation of predicting score distributions lies in the possibility to derive aesthetic subjectivity [5], [6], [8]. Therefore, we first consider state-of-the-art aesthetic distribution predictors (see Section II) to estimate the subjectivity measures introduced above, as illustrated in Figure 3(a). The advantage of this approach is that, once the distribution is estimated, one can compute any subjectivity measure from it. However, we will show experimentally that this approach is generally sub-optimal compared to directly estimating a subjectivity score.



(a) Indirect subjectivity prediction



(b) Direct subjectivity prediction

Fig. 3. Subjectivity Prediction Framework. In the indirect prediction framework, an aesthetic score distribution is estimated first, and subjectivity measures are computed over it. We compare this approach with directly predicting subjectivity computed on ground-truth distributions (b).

Direct subjectivity prediction: A limitation of the indirect subjectivity prediction is that the estimated distribution scores are generally noisy, as even the best method to predict the histogram of aesthetic scores has limited performance [8]. In principle, assuming a deep neural network predicting aesthetic distribution approximates the maximum likelihood estimator [21], a subjectivity estimator based on the predicted distributions is asymptotically efficient [22]. However, in practice the number of samples used for training the network is finite, and prediction errors on the distribution might lead to worse prediction performance of subjectivity.

Therefore, we consider the alternative approach consisting in predicting directly the ground-truth subjectivity measures, as shown in Figure 3(b). The subjectivity measures are computed on the ground-truth aesthetic distribution. We use afterwards a deep convolutional neural network to predict these subjectivity measures.

IV. Experimental results

In this section we analyze the prediction performance of the aesthetic subjectivity measures introduced in Section III.

A. Experimental setup

We choose Resnet-34 as network structure to predict subjectivity. According to our experiments and the test in [9], Resnet-34 provides similar accuracy result as VGG-16, but uses less memory. In addition, we also use Resnet-101 to study the influence of a deeper network structure in the direct aesthetic subjectivity prediction. The last (fully connected) layer of Resnet-34 is replaced by 3 fully connected layers: two 512×512 fully connected layers plus a 512×1 layer to give a 1-dimensional output. For Resnet-101, the two additional fully connected layers have size 2048×2048. The drop out rate is 0.5 for every fully connected layer.

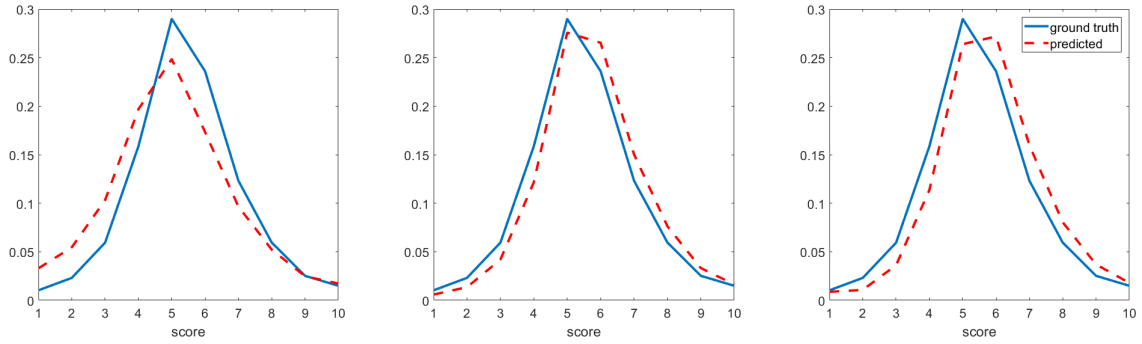
We use Pytorch models pre-trained on ImageNet [23], and we fine-tune them using training images from the AVA dataset [2]. We use the standard test set of AVA as in previous work, which consists of 19,930 images. This leaves 260,264 images for train and validation. We randomly pick 23,553 pictures for validation, corresponding to approximately 10% of the training set size. All of the input images are resized to 224×224 pixels. Even though previous methods often augment data with horizontal flip, we decide not to do any kind of data augmentation, as differently from classification or recognition tasks, the ground truth in aesthetics is obtained by human raters and might be influenced by flipping. We employ Adam optimizer [24] and a batch size of 64. The learning rate is decreased by 10 times when the loss does not change over two consecutive epochs. We fix the initial learning rate to 10^{-5} and the maximum number of iterations to 40,000. We employ the L1 norm as loss function for the direct prediction of the subjectivity measures.

For the indirect subjectivity prediction, we consider the following three methods for predicting aesthetic score distributions: the work of Bin Jin et al. [5] (chi-square distance loss); NIMA [7] (Earth Mover's Distance loss); and the RSCJS method of Jin et al. [8] (cumulative Jensen-Shannon divergence loss). Bin Jin et al. provide their trained model (using VGG-16), but use a different (smaller) test set than the standard AVA one. Since their test set is not provided, and for the sake of a fair comparison with other methods, We run their model on the standard AVA test set instead. For NIMA and RSCJS, the original code is not available, and we reimplemented them following the original papers. For NIMA and RSCJS, we use Resnet-34, modified as discussed above.

B. Performance Indicators

We evaluate the prediction of the subjectivity measures using 4 performance indicators:

- *Pearson's Linear Correlation Coefficient* (PLCC), which measures the linearity of the relationship between the predicted and the ground-truth subjectivity score. Higher values indicate better prediction performance.
- *Spearman's Rank-Order Correlation Coefficient* (SROCC), which indicates the degree of monotonicity of the prediction. Higher values indicate better prediction performance.
- *Mean absolute error* (MAE), which indicates the degree of accuracy. Prediction is more accurate when MAE is small.



(a) Bin Jin's (b) NIMA (c) RSCJS

Fig. 4. Average predicted score distribution vs. ground-truth score distribution over the test set, for the three considered state-of-the-art distribution prediction methods. Notice that for all of them, the average predicted distribution is shifted compared to the original.

TABLE I
PEARSON'S LINEAR CORRELATION COEFFICIENT (PLCC)

Methods	*SD*	*MAD*	*MED*	*DUD*
Bin Jin's [5]	0.145	0.159	0.178	0.096
NIMA [7]	0.169	0.187	0.211	0.255
RSCJS [8]	0.187	0.199	0.227	0.281
Direct (Resnet-34)	0.274	0.276	0.323	0.351
Direct (Resnet-101)	**0.307**	**0.304**	**0.333**	**0.360**

TABLE II
SPEARMAN'S RANK-ORDER CORRELATION COEFFICIENT (SROCC)

Methods	*SD*	*MAD*	*MED*	*DUD*
Bin Jin's [5]	0.142	0.156	0.162	0.085
NIMA [7]	0.230	0.240	0.250	0.297
RSCJS [8]	0.169	0.152	0.228	0.283
Direct (Resnet-34)	0.267	0.268	0.311	0.351
Direct (Resnet-101)	**0.295**	**0.295**	**0.316**	**0.355**

TABLE III
MEAN ABSOLUTE ERROR (MAE)

Methods	*SD*	*MAD*	*MED*	*DUD*
Bin Jin's [5]	0.294	0.255	0.106	0.226
NIMA [7]	0.171	0.156	0.059	0.122
RSCJS [8]	0.169	0.152	0.059	0.117
Direct (Resnet-34)	0.148	0.133	0.054	0.120
Direct (Resnet-101)	**0.146**	**0.132**	**0.053**	**0.101**

TABLE IV
MEAN RELATIVE ABSOLUTE ERROR (MRAE)

Methods	*SD*	*MAD*	*MED*	*DUD*
Bin Jin's [5]	0.226	0.270	0.210	0.053
NIMA [7]	0.129	0.162	0.128	0.029
RSCJS [8]	0.127	0.158	0.127	0.028
Direct (Resnet-34)	0.107	**0.130**	0.126	0.025
Direct (Resnet-101)	**0.104**	**0.130**	**0.120**	**0.024**

- *Mean relative absolute error* (MRAE), which is MAE normalized by ground-truth values. A smaller MRAE indicates higher accuracy.

C. Experiment Results

Tables I-IV show direct and indirect subjectivity prediction performance. We observe that direct subjectivity prediction always outperforms indirect prediction through distribution scores, for all the proposed subjectivity measures. In particular, for the same network complexity (Resnet-34), predicting directly the subjectivity is clearly better than predicting the score distribution first and computing subjectivity based on it. A possible explanation can be obtained by looking at the results of distribution prediction, as shown in Figure 4, which compares the average predicted aesthetic score distribution vs. the average ground-truth one. For the three distribution prediction methods considered here, we notice that, on average, predicted distributions are different from the original and may even be shifted. Notice that all of the proposed subjectivity measures are affected by errors in the prediction of the histogram.

Although direct prediction improves all the considered performance indicators, we observe that overall the prediction performance is still not satisfactory, e.g., the SROCC is just slightly above 0.4. We might wonder whether this is due to a limited capacity of the Resnet-34 model we employed. Therefore, in order to study how subjectivity prediction performance improves with a more complex network, we tested the direct prediction scheme using Resnet-101, which is much deeper than Resnet-34. As expected, the results generally improve over the simpler Resnet-34. However, this improvement is in most case only marginal, showing that aesthetic subjectivity prediction is intrinsically a hard problem – at least a harder one than predicting the average aesthetic score, where SROCC between predicted and ground-truth values is higher than 0.6 [7].

Comparing the different subjectivity measures, those inspired by information theory (DUD and MED) are in general those with higher prediction performance. Among the statistical motivated descriptors, the SD is generally predicted more accurately than MAD. We can assume that, for the same neural

network model complexity, a ground-truth variable which has a higher dependence on the input is easier to predict, or, in other terms, target variables which tend to be more "noisy" will be more difficult to learn. Thus, we can argue that the subjectivity measures based on information theory are somewhat more robust than statistical deviation measures. A possible rationale behind this could be that both DUD and MED are based on distances between histograms, which take into account the whole score distribution. On the other hand, SD completely captures data variability when the underlying score distribution is Gaussian, which is the case for only 62% of AVA images [8]. MAD is supposed to be more robust to skewed distributions, but it might be affected by the sample median computation, which on a 10-dimensional distribution as for aesthetic scores can only take values over a small set, i.e., $\{1, 1.5, 2, \ldots, 10\}$.

Notice that the DUD measure achieves the best correlation among the four subjectivity measures, despite the fact that it penalizes more those images with distributions having mean score far from the midpoint of the quality scale. These are also the images that which are less frequent in the AVA dataset. Therefore, DUD might implicitly act as a weighting scheme during learning, similar to [5]. However, this effect might be less evident on more balanced datasets, and should be verified by further experimental evidence. We leave this to future work.

V. Conclusion

In this paper we have analyzed the problem of defining and predicting aesthetic score subjectivity, intended as the degree of consensus human raters express about the aesthetic value of a picture. To this end, we have considered several measures of subjectivity, and two possible subjectivity prediction frameworks.

Among the analyzed descriptors of subjectivity, we have found that our newly proposed measures inspired by information theoretical principles are, in general, easier to learn, indicating that they might be more discriminative and robust compared to simpler statistical deviation measures.

We have shown that predicting subjectivity from predicted score distributions is, in general, sub-optimal compared to directly predicting it from ground-truth subjectivity scores. This indicates that, in practice, aesthetic score distribution predictors are not sufficiently accurate to enable assessing correctly the aesthetic subjectivity.

Despite our approach achieves state-of-the-art subjectivity prediction performance, we recognize that predicting subjectivity is a much harder task than predicting, e.g., the average aesthetic score – histogram-prediction-based methods can achieve correlations of 0.6 or higher for that task. We believe that this is partially due, in addition to the complexity of the task in itself, to the noisy nature of current aesthetic datasets. This is evident for the benchmark AVA dataset, where aesthetic scores are influenced by many factors that go beyond the pure aesthetic value of a picture. Building cleaner and more reliable aesthetic datasets is among the future directions to consider in this field.

References

[1] Y. Deng, C. C. Loy, and X. Tang, "Image aesthetic assessment: An experimental survey," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 80–106, 2017.

[2] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: a large-scale database for aesthetic visual analysis," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2408–2415.

[3] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *European Conference on Computer Vision*. Springer, 2016, pp. 662–679.

[4] R. Reber, N. Schwarz, and P. Winkielman, "Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?" *Personality and social psychology review*, vol. 8, no. 4, pp. 364–382, 2004.

[5] B. Jin, M. V. O. Segovia, and S. Süsstrunk, "Image aesthetic predictors based on weighted CNNs," in *IEEE International Conference on Image Processing*. Phoenix, AZ, USA: Ieee, October 2016, pp. 2291–2295.

[6] N. Murray and A. Gordo, "A deep architecture for unified aesthetic prediction," *arXiv preprint arXiv:1708.04890*, 2017.

[7] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3998–4011, 2018.

[8] X. Jin, L. Wu, X. Li, S. Chen, S. Peng, J. Chi, S. Ge, C. Song, and G. Zhao, "Predicting aesthetic score distribution through cumulative jensen-shannon divergence," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[9] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, pp. 64 270–64 277, 2018.

[10] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *European conference on computer vision*. Springer, 2006, pp. 288–301.

[11] T. O. Aydın, A. Smolic, and M. Gross, "Automated aesthetic analysis of photographic images," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 1, pp. 31–42, 2014.

[12] V. Hulusic, G. Valenzise, E. Provenzi, K. Debattista, and F. Dufaux, "Perceived dynamic range of HDR images," in *IEEE Int. Conference on Quality of Multimedia Experience*, 2016, pp. 1–6.

[13] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rapid: Rating pictorial aesthetics using deep learning," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 457–466.

[14] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 990–998.

[15] K. Park, S. Hong, M. Baek, and B. Han, "Personalized image aesthetic quality assessment by joint regression and ranking," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 1206–1214.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[18] W. Luo, X. Wang, and X. Tang, "Content-based photo quality assessment," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 2206–2213.

[19] X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 1930–1943, 2013.

[20] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[22] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.