IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 November 2023, accepted 2 December 2023, date of publication 22 December 2023, date of current version 29 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3345921

RESEARCH ARTICLE

Probabilistic Modeling of Image Aesthetic Assessment Toward Measuring Subjectivity

HYEONGNAM JANG[1], **YEEJIN LEE**[2], **AND JONG-SEOK LEE**[1], **(Senior Member, IEEE)**
[1]School of Integrated Technology, Yonsei University, Incheon 21983, South Korea
[2]Department of Electrical and Information Engineering, Seoul National University of Science and Technology, Seoul 01811, South Korea

Corresponding author: Jong-Seok Lee (jong-seok.lee@yonsei.ac.kr)

This work was supported in part by the Artificial Intelligence Graduate School Program, Yonsei University, under Grant 2020-0-01361; and in part by NRF funded by the Korean Government [Ministry of Science and ICT (MSIT)] under Grant 2021R1A2C2011474.

ABSTRACT Assessing image aesthetics is a challenging computer vision task. One reason is that aesthetic preference is highly subjective and may vary significantly among people for certain images. Thus, it is important to properly model and quantify such *subjectivity*, but there has not been much effort to resolve this issue. In this paper, we propose a novel probabilistic framework that can model and quantify subjective aesthetic preference based on the subjective logic. In this framework, the rating distribution is modeled as a beta distribution, from which the probabilities of being definitely pleasing, being definitely unpleasing, and being uncertain can be obtained. We use the probability of being uncertain to define an intuitive metric of subjectivity. Furthermore, we present a method to learn deep neural networks for prediction of image aesthetics, which is shown to be effective in improving the performance of subjectivity prediction via experiments.

INDEX TERMS Deep learning, image aesthetic assessment, subjectivity, subjective logic, aesthetic uncertainty.

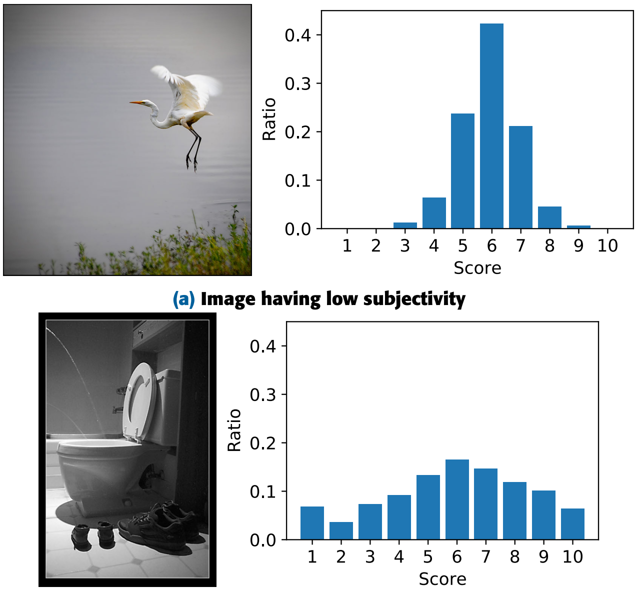
I. INTRODUCTION

Image aesthetic assessment is to automatically evaluate the image in the aesthetic viewpoint, i.e., how aesthetically pleasing to human viewers an image will be. It is a challenging computer vision task since it requires to imitate high-level aesthetic perception of humans, but it can be useful in many applications including image search and retrieval, recommender systems, image enhancement, etc.

In order to obtain the aesthetic ground truth of an image, it is usual to ask a group of raters to provide aesthetic scores for the image. In the binary classification task, the image is considered as aesthetically pleasing if the mean score is higher than a threshold, and as unpleasing otherwise. It is also possible to set a regression task where the mean rating score is predicted. However, these tasks do not consider the diversity in the raters' opinions. In Fig. 1, two example images and their distributions of the rating scores from raters are shown, which are from the AVA dataset [1]. The two images have similar mean scores (about 5.92), so that they have the same target mean rating score for mean rating regression and the same class label for binary classification. However, these do not capture the different levels of diversity in the raters' opinions between the two cases. While the distribution of ratings in Fig. 1a is concentrated around the mean score, that in Fig. 1b is spread widely over the whole score range. Thus, in the case of Fig. 1b, the results of the mean rating regression and binary classification may be disagreed by a significant proportion of users.

Therefore, it is necessary to consider *subjectivity* in aesthetic assessment. In the image aesthetic assessment, the subjectivity can be said to be the degree to which people's aesthetic evaluations of an image differ. Understanding and predicting aesthetic subjectivity can be beneficial in practical applications. For instance, in an image recommendation system considering aesthetics, images that are expected to

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.



(a) **Image having low subjectivity**

(b) **Image having high subjectivity**

FIGURE 1. Example images and their distributions of the rating scores given by raters. The two images have similar mean rating scores (5.9167 and 5.9174 respectively) but different degrees of subjectivity.

have high subjectivity could be excluded from the set of recommended images in order to improve general users' satisfaction with the recommendation. However, only a limited number of existing studies have dealt with the issue of subjectivity. The most widely used way to measure the subjectivity is to compute the standard deviation (STD) (or variance) of the rating scores given by multiple viewers. However, it has an issue in terms of interpretability because the meaning of its scale is not sufficiently intuitive. Similar metrics exist, such as mean absolute deviation around median [2], which are suffer from the same problem.

In this paper, we propose a novel probabilistic framework for modeling and quantifying the subjectivity of image aesthetics based on the subjective logic [3]. In this framework, the rating distribution of an image is modeled as a beta distribution, from which the probabilities of being definitely pleasing, being definitely unpleasing, and being uncertain can be obtained simultaneously. In particular, the probability of being uncertain defines an intuitive metric of subjectivity, named *aesthetic uncertainty*. Unlike the existing subjectivity metrics, it is a probability measure, which can be easily interpreted. Furthermore, we present a method for predicting image aesthetics by modeling the rating distribution as the beta distribution. In addition, it is demonstrated that users' satisfaction can be enhanced by predicting whether the aesthetic level of an image is uncertain due to high subjectivity.

The paper is organized as follows. The related work is surveyed in Section II. In Section III, the subjectivity modeling of image aesthetics is described. Section IV presents our method to predict the image aesthetics by the subjectivity modeling. Section V presents the experiments to evaluate the performance of our method. In Section VI, the results with analysis are shown. Finally, conclusions are given in Section VII.

II. RELATED WORK

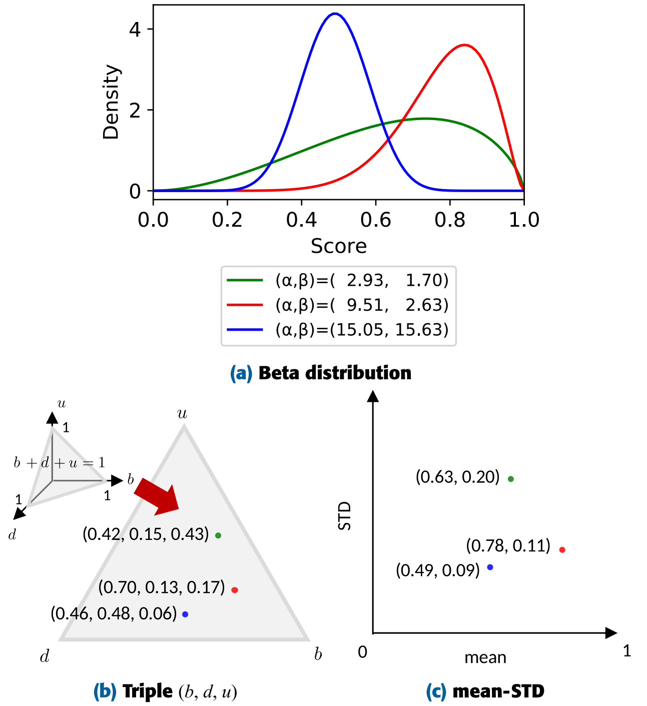
A. IMAGE AESTHETIC ASSESSMENT

In literature, three tasks have been mainly considered for automatic aesthetic assessment of images: binary classification, mean score regression, and rating distribution prediction. The binary classification task distinguishing pleasing vs. unpleasing images has been considered most popularly. There exist several methods, from those using handcrafted features [4], [5], [6] to deep learning-based methods [7], [8], [9], [10]. To obtain more informative results than binary class information, the mean rating regression task has been addressed [11], [12], [13]. Prediction of the whole rating distribution has been also considered [14], [15], [16], [17], [18], [19], [20], which is the most challenging but has potential to provide the most comprehensive information regarding the aesthetic characteristics of the given image. In this paper, we consider all these tasks and also the subjectivity regression task.

B. SUBJECTIVITY OF IMAGE AESTHETICS

Subjectivity is a clearly distinguished issue in image aesthetics compared to other image-based problems such as object classification. While the class of an object in an image can be objectively determined, subjective judgement is involved in image aesthetics, and thus an image preferred by certain viewers is not necessarily pleasing to some other viewers. Park and Zhang [21] modeled the human aesthetic evaluation process by a dynamic system and showed that the response time for aesthetic evaluation of an image is related to the subjectivity level of the image in terms of STD. Kim et al. [22] analyzed the relationship between subjectivity (expressed by STD) and user comments, which showed that several factors such as unusualness and coexistence of aesthetic merits and demerits are involved in determining the level of subjectivity of an image. There also exist studies suggesting personalized image aesthetic assessment techniques that reflect the information of a specific user [12], [23], [24], [25], [26], [27], [28].

The rating distribution itself can give the information regarding subjectivity (as in Fig. 1) but only implicitly [2]. Therefore, there is a need to explicitly quantify subjectivity as a scalar value. However, there is not much progress in this research direction. STD has been mostly used [14], [29]. Kang et al. [2] defined additional subjectivity metrics including the mean absolute deviation around the median (MAD), distance to uniform distribution (DUD), and distance from the maximum entropy distribution (MED). However, all these metrics have a limitation in interpretability because they have neither upper limits nor interpretable scales (except 0). Our work addresses this issue and proposes an intuitive metric of subjectivity.



(a) **Beta distribution**

(b) **Triple** (b, d, u)

(c) **mean-STD**

FIGURE 2. Image aesthetics via the subjectivity modeling. The fitted beta distributions and the corresponding triples of (b, d, u) on the equilateral triangle are shown in (a) and (b), respectively. For comparison, the means and STDs of theratting distributions are shown in (c).

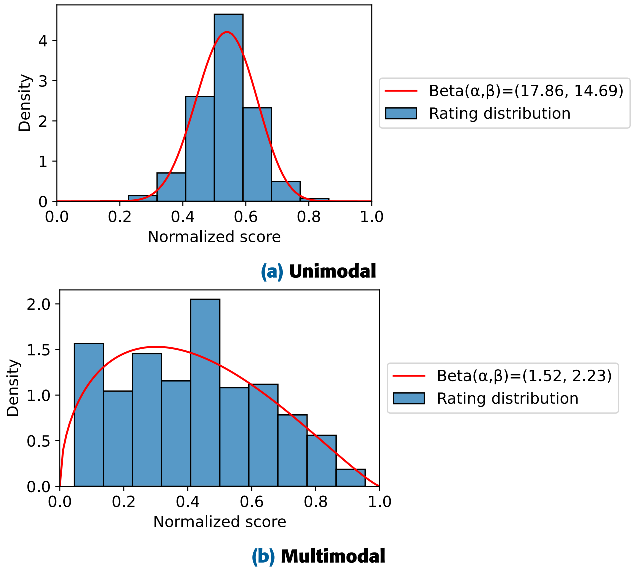
III. PROPOSED METHOD FOR SUBJECTIVITY MODELING

The proposed framework is based on the subjective logic [3]. The subjective logic is a probabilistic reasoning for modeling subjective opinions involving uncertainty. An aesthetic rating chosen from a range of scores (e.g., 1 to 10) is a multinomial opinion. However, if the rating scale is normalized between 0 and 1, the rating can be considered as a probability of aesthetic pleasingness in the binary classification [11], i.e., a binomial opinion indicating the subjective belief about pleasingness. Then, the opinion is represented by three components [3]: b (belief mass), d (disbelief mass), and u (uncertainty mass), which correspond to the probabilities of being definitely pleasing, being definitely unpleasing, and being uncertain, respectively. These components have values between 0 and 1, and satisfy $b + d + u = 1$. As a result, the aesthetics of an image can be represented as a point on an equilateral triangle as shown in Fig. 2b.

The binomial opinion can be modeled by a beta distribution whose probability density function (PDF) is given by

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \tag{1}$$

where $0 \le x \le 1$, α and β are shape parameters, and $B(\alpha, \beta)$ is a normalization constant ensuring $\int_0^1 f(x; \alpha, \beta)dx = 1$. When the rating distribution of an image is given, a beta distribution is fitted to the distribution by finding the optimal values of α and β that minimize the difference (e.g., earth movers' distance (EMD)) between the given and fitted distributions.



(a) **Unimodal**

(b) **Multimodal**

FIGURE 3. Examples of unimodal and multimodal aesthetic rating distributions (after score normalization) and the fitted beta distributions.

Finally, the three probabilities (b, d, and u) can be obtained from the fitted values of α and β as follows [3]:

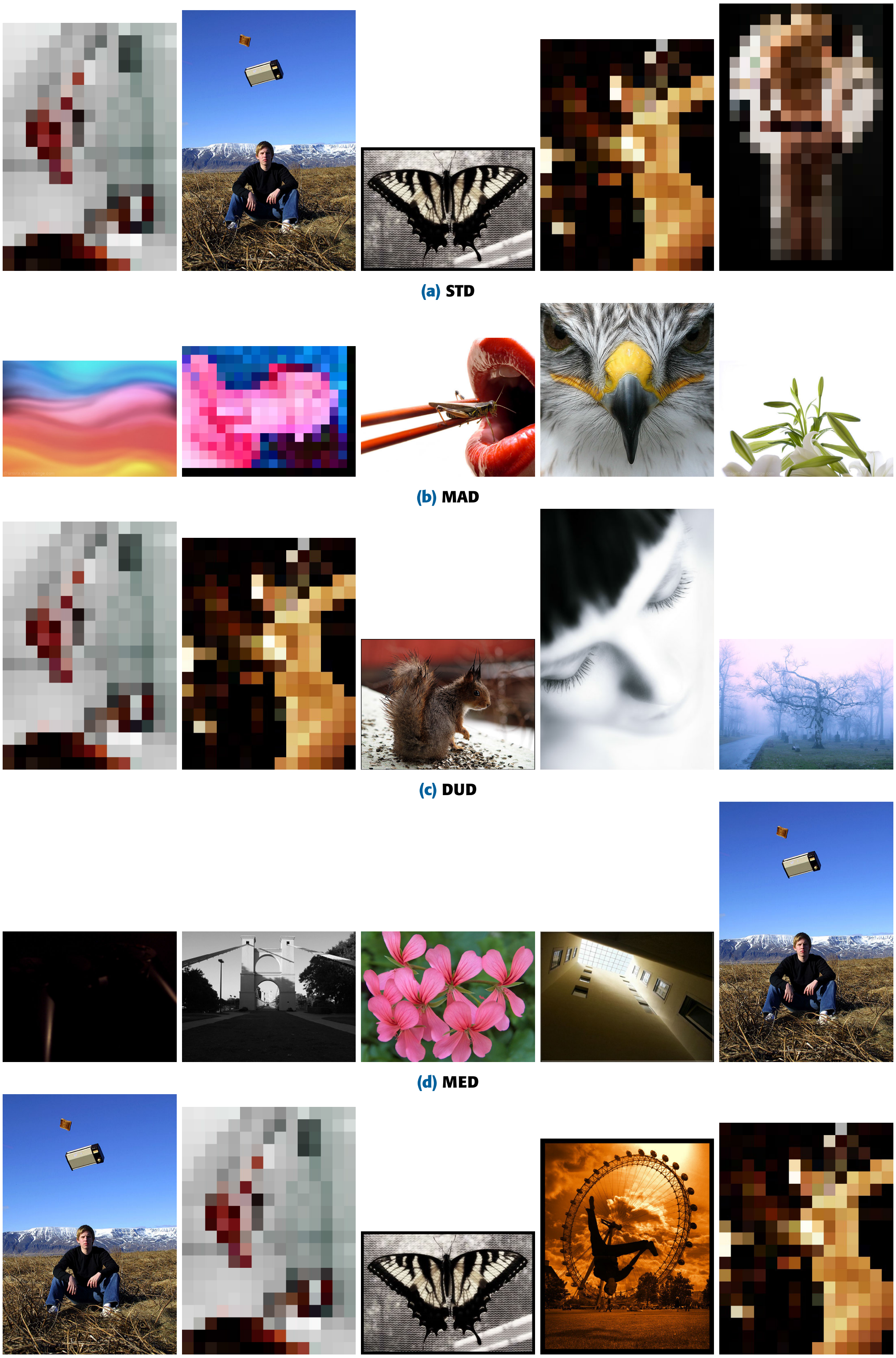
$$b = \frac{\alpha - 1}{\alpha + \beta}, \quad d = \frac{\beta - 1}{\alpha + \beta}, \quad u = \frac{2}{\alpha + \beta}. \tag{2}$$

Fig. 2 shows examples of fitted beta distributions and their representations in the equilateral triangle of b, d, and u. The image corresponding to the red-colored distribution in Fig. 2a would be considered as aesthetically pleasing by most people, which is reflected in the large value of b and the small value of u in Fig. 2b. The blue-colored case in Fig. 2a is judged to have an intermediate level of aesthetics without much disagreement among people, which is represented by the small value of u and the similar values of b and d in Fig. 2b. The green-colored case would be classified as pleasing if binary classification is performed, but it involves high subjectivity; thus, u appears to be large while $b > d$.

The beta distribution is unimodal, while an original rating distribution may be multimodal. An example case is shown in Fig. 3b, along with the fitted beta distribution. For the images in the AVA dataset [1] used in our experiments, we conduct the dip test of unimodality [30] and find that 94.56% of the images have unimodal distributions, whereas only 5.14% and 0.30% are bimodal and trimodal, respectively. Note that some of the multimodal distributions may be due to noise in the ratings, which can be reduced by the unimodal modeling. Thus, we can say that fitting to beta distributions is reasonable.

A. AESTHETIC UNCERTAINTY (AESU)

While the rating distribution itself contains the information of subjectivity, it is practically useful to obtain a single-valued metric quantifying the level of subjectivity from the distribution. For this, we define a new metric called aesthetic



(a) STD

(b) MAD

(c) DUD

(d) MED

(e) AesU

FIGURE 4. Five images showing the highest subjectivity determined by each subjectivity measure in the descending order of subjectivity. Suggestive images are pixelated.

uncertainty (AesU), by u obtained from the subjectivity modeling. It has several advantages compared to the existing metrics. Since AesU is a probability within [0, 1], one can intuitively grasp the level of subjectivity, whereas the scales of the existing metrics (such as STD, MAD, DUD, and MED) are not sufficiently intuitive. In addition, AesU can be

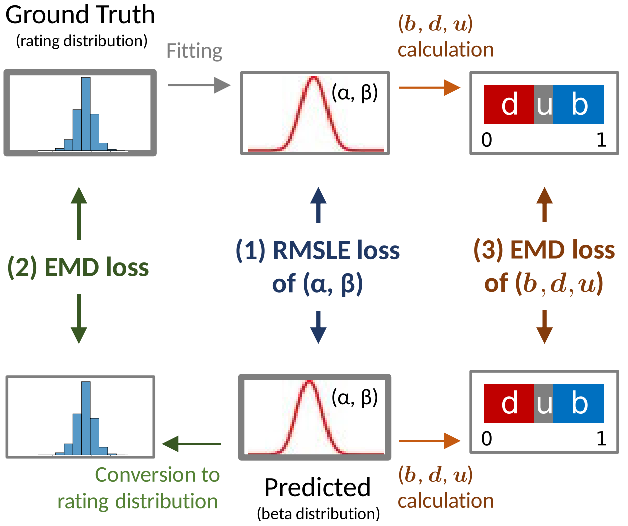


FIGURE 5. Ways of measuring the loss of the predicted beta distribution. (1) The predicted shape parameters are converted to a rating distribution, which is compared to the ground truth rating distribution in terms of EMD. (2) RMSLE is measured between the predicted shape parameters ($\alpha_{pred}, \beta_{pred}$) and the ground truth shape parameters (α_{GT}, β_{GT}) (obtained by fitting to the ground truth rating distribution). (3) EMD is measured between the triplet (b, d, u) calculated from ($\alpha_{pred}, \beta_{pred}$) and that from ($\alpha_{GT}, \beta_{GT}$).

interpreted together with the other two probabilities (b and u) as discussed in Fig. 2b, which is difficult with the mean and STD (or, MAD, DUD, MED) of ratings.

Fig. 4, shows five images with the highest subjectivity determined by each subjectivity measure among the test set of the AVA dataset. Most of these images are suggestive or atypical in terms of contents, perspective, composition, or contrast. It is observed that STD and AesU show similar results with four common images, whereas zero, two and one images are common between STD and each of MAD, DUD, and MED, respectively. Additionally, the Pearson correlation coefficient between STD and AesU for the whole test dataset reaches 0.91. Thus, it can be seen that AesU has reliability comparable to STD.

IV. PREDICTING IMAGE AESTHETICS WITH BETA DISTRIBUTION

Based on our aesthetic modeling framework, we propose a method to train a neural network model for prediction of image aesthetics. Unlike existing methods [14], [15], [16], [17], we do not need to obtain a predicted rating distribution because the ground truth rating distribution is modeled by the beta distribution as in the previous section, which can be fully described by α and β. Therefore, it is sufficient to make the neural network predict α and β of the beta distribution. Then, all tasks, including prediction of subjectivity measures as well as binary classification, mean rating regression, and rating distribution prediction, can be performed.

For training of the model, we design three candidates for the loss function, which are illustrated in Fig. 5. First, the root mean squared log error (RMSLE) between the shape parameters of the beta distribution fitted to the ground truth rating distribution (α_{GT}, β_{GT}), and those predicted by the model, ($\alpha_{pred}, \beta_{pred}$). Here, RMSLE is used instead of mean squared error (MSE) or root mean squared error (RMSE) in order to effectively handle the shape parameters that become larger by orders of magnitude occasionally for some images. Second, the predicted shape parameters are converted to the corresponding rating distribution, which is compared to the ground truth rating distribution in terms of EMD. Finally, the three probabilities (b, d and u), which are easily calculated by Eq. (2), can be used. In other words, EMD between (b, d, u) from the predicted beta distribution and that of the ground truth beta distribution is used as the loss. We compare these candidates experimentally, from which we decide to use the third one (see Section VI-A).

V. EXPERIMENTS

A. DATASET

We use the AVA dataset [1]. It contains photos and their aesthetic ratings from challenges of DPChallenge.[1] The dataset is composed of 256,000 images. 236,000 images are for training and 20,000 are for test.

B. FITTING RATING TO BETA DISTRIBUTION

Before the training step, we convert the rating distribution to beta distribution. We use an optimization function to minimize EMD between the original rating distribution and the fitted beta distribution.

C. BACKBONE MODELS

We use popular generic CNN models and a latest image aesthetic assessment model as backbone models for our experiments. The former includes VGG16 [31], ResNet-50 [32], and ConvNeXT [33], and MaxViT [18], and the latter corresponds to the hierarchical layout-aware graph convolutional network (HLAGCN) [17]. HLAGCN models the complex relations among interesting regions in the input image using a graph convolutional network. MaxViT is a transformer model that considers the global context by using a multi-axis attention. We use the tiny structure of ConvNeXT and MaxViT (noted as '-T') because of the limitation of the computing power.

D. APPROACHES

A model produces two output values (the shape parameters of a beta distribution, i.e., α and β). For comparison, we employ the conventional approach that directly predicts the rating distribution [14], [15], [16], [17].

E. PERFORMANCE MEASURES

We consider various tasks of image aesthetic assessment, and use appropriate performance measures for each task by

[1] http://www.dpchallenge.com

TABLE 1. Results of subjectivity regression in terms of (a) PLCC, (b) MAE.

Backbone	Approach	STD	MAD	MED	DUD	AesU
VGG16	Conventional (rating distribution)	0.2390	0.2391	0.3041	0.3273	0.2286
	Proposed (beta distribution)	0.2526	0.2437	0.3072	0.3358	0.2478
ResNet50	Conventional (rating distribution)	0.2897	0.2755	0.3647	0.4111	0.2783
	Proposed (beta distribution)	0.2688	0.2538	0.3356	0.3959	0.2647
HLAGCN	Conventional (rating distribution)	0.2801	0.2661	0.3538	0.3538	0.2693
	Proposed (beta distribution)	0.2913	0.2817	0.3600	0.3907	0.2902
ConvNeXT-T	Conventional (rating distribution)	0.3417	0.3291	0.4046	0.4465	0.3181
	Proposed (beta distribution)	0.3499	0.3504	0.4129	0.4558	0.3538
MaxViT-T	Conventional (rating distribution)	0.3246	0.3097	0.3931	0.4457	0.3060
	Proposed (beta distribution)	0.3302	0.3254	0.3966	0.4435	0.3327

(a) PLCC as the performance measure. The higher the value is, the better the performance is.

Backbone	Approach	STD	MAD	MED	DUD	AesU
VGG16	Conventional (rating distribution)	0.1649	0.1587	0.0541	0.0609	0.0351
	Proposed (beta distribution)	0.1783	0.1504	0.0555	0.0561	0.0328
ResNet50	Conventional (rating distribution)	0.1485	0.1472	0.0526	0.0572	0.0346
	Proposed (beta distribution)	0.1900	0.1520	0.0566	0.0529	0.0323
HLAGCN	Conventional (rating distribution)	0.1536	0.1508	0.0532	0.0619	0.0344
	Proposed (beta distribution)	0.1852	0.1503	0.0556	0.0531	0.0320
ConvNeXT-T	Conventional (rating distribution)	0.1452	0.1436	0.0515	0.0551	0.0343
	Proposed (beta distribution)	0.1544	0.1417	0.0515	0.0550	0.0330
MaxViT-T	Conventional (rating distribution)	0.1480	0.1454	0.0519	0.0564	0.0339
	Proposed (beta distribution)	0.1788	0.1463	0.0537	0.0517	0.0314

(a) MAE as the performance measure. The lower the value is, the better the performance is.

TABLE 2. Results of binary classification, mean rating regression, and rating distribution prediction. The accuracy is used as the performance measure of the binary classification. PLCC and MAE are used as the performance measure of the mean rating regression. EMD is used for the rating distribution prediction.

Backbone	Approach	Binary classification	Mean rating		Distribution
		Accuracy ↑	PLCC ↑	MAE ↓	EMD ↓
VGG16	Conventional (rating distribution)	0.7781	0.6559	0.4509	0.0504
	Proposed (beta distribution)	0.7843	0.6494	0.4484	0.0507
ResNet50	Conventional (rating distribution)	0.8057	0.7166	0.4099	0.0448
	Proposed (beta distribution)	0.7990	0.7041	0.4199	0.0481
HLAGCN	Conventional (rating distribution)	0.7958	0.7049	0.4184	0.0458
	Proposed (beta distribution)	0.8011	0.7103	0.4142	0.0475
ConvNeXT-T	Conventional (rating distribution)	0.8171	0.7472	0.3876	0.0426
	Proposed (beta distribution)	0.8188	0.7527	0.3845	0.0447
MaxViT-T	Conventional (rating distribution)	0.8173	0.7475	0.3870	0.0426
	Proposed (beta distribution)	0.8114	0.7437	0.3907	0.0454

following the previous studies [2], [14]. Primarily, we conduct the subjectivity prediction task, where we use STD, MAD, DUD, MED, and AesU as the subjectivity measure. We also consider the widely used image aesthetic assessment tasks: binary classification, mean score regression, and rating distribution prediction.

Our method produces a predicted beta distribution (in the form of its shape parameters α and β). Except AesU,

TABLE 3. Comparison of the ways to measure the loss (see Fig. 5).

Loss	STD		MAD		MED		DUD		AesU	
	PLCC ↑	MAE ↓	PLCC ↑	MAE ↓	PLCC ↑	MAE ↓	PLCC ↑	MAE ↓	PLCC ↑	MAE ↓
(1) EMD of rating distribution	0.3417	0.1452	0.3291	0.1436	0.4046	0.0515	0.4465	0.0551	0.3181	0.0343
(2) RMSLE of Beta (α, β)	0.3595	0.1531	0.3483	0.1417	0.4053	0.0514	0.3911	0.0586	0.3558	0.0329
(3) EMD of Beta (b, d, u)	0.3499	0.1544	0.3504	0.1417	0.4129	0.0515	0.4558	0.0550	0.3538	0.0330

(a) Subjectivity regression

Loss	Binary classification	Mean rating		Distribution
	Accuracy ↑	PLCC ↑	MAE ↓	EMD ↓
(1) EMD of rating distribution	0.8171	0.7472	0.3876	0.0426
(2) RMSLE of Beta (α, β)	0.7933	0.6901	0.4288	0.0486
(3) EMD of Beta (b, d, u)	0.8188	0.7527	0.3845	0.0447

(b) Binary classification, mean rating regression, and rating distribution prediction

which is directly available from the shape parameters using Eq. (2), the subjectivity measures are obtained from the rating distribution converted from the predicted beta distribution or the ground truth rating distribution.

For the binary classification task, the classification accuracy is used. The threshold of the two classes is set to a score of 5 in the scale of 1 to 10 as in [7], [8], [9], and [10]. The performance of mean rating regression is measured in terms of Pearson linear correlation coefficient (PLCC), and mean absolute error (MAE) between the ground truth mean ratings and the predicted mean ratings. The same measures are also used for the subjectivity regression task. For distribution prediction, we use EMD between the ground truth and predicted distributions.

F. IMPLEMENTATION DETAILS

All experiments are conducted on a PC that has AMD Ryzen 5 5600X CPU, 128GB of RAM, NVIDIA Geforce RTX 3090 24GB GPU, and Microsoft Windows 10. We use Python 3.9.7, PyTorch 1.10.1, CUDA 11.3, and cuDNN 8.0.

We divide the training dataset further into 223,000 training images and 13,000 validation images. The images are resized to 256 × 256 and randomly cropped to 224 × 224. and a random horizontal flip is applied for data augmentation.

To train the models, we use the SGD optimization with a batch size of 48, a Nesterov momentum parameter of 0.9 and a weight decay parameter of 5×10^{-4}. The learning rate is reduced by 5% every 10 epochs. The models are trained for 100 epochs, but if the validation loss does not decrease for 30 epochs, the learning is stopped.

VI. RESULTS

Table 1 shows the results of our main task, i.e., subjectivity regression, for STD, MAD, MED, DUD, and AesU. The proposed approach shows improved performance compared to the conventional approach in most cases. When the backbone models are compared, ConvNeXT-T performs the best, and MaxViT-T also performs well. For ConvNeXT-T, the PLCC improvement by the proposed approach over the conventional one is 4.8% on average, with the maximum improvement by 11.2% for AesU. These results demonstrate that the proposed approach using the beta distribution is effective to learn the subjectivity of image aesthetics. In particular, the proposed subjectivity modeling plays a key role by reducing the noise in the raw rating distribution.

We additionally show the results of the other tasks, i.e., binary classification, mean rating regression, and rating distribution prediction, in Table 2. The proposed approach shows the performance comparable to that of the conventional approach. In particular, the proposed approach for ConvNeXT-T shows the best performance among various experimental conditions except the rating distribution prediction. The performance of the rating distribution prediction is lower for the proposed approach than the conventional approach in all cases, which appears to be due to the error in approximation with the beta distribution.

A. COMPARISON OF LOSSES

We conduct experiments to compare the three ways to measure the loss, which are explained in Section IV. Table 3 shows the comparison results for ConvNeXT-T. For the subjectivity regression (Table 3a), EMD of rating distribution shows the lowest performance, while the other two are comparable. For the other tasks (Table 3b), RMSLE of (α, β) performs worse than the other two showing similar performance. Overall, EMD of (b, d, u) shows the best performance. Through this, we can see that AesU, along with b and d, expresses aesthetic rating characteristics better than the rating distribution for model training.

B. APPLICATION OF AESU

We illustrate an application scenario where our subjectivity modeling can be beneficial.

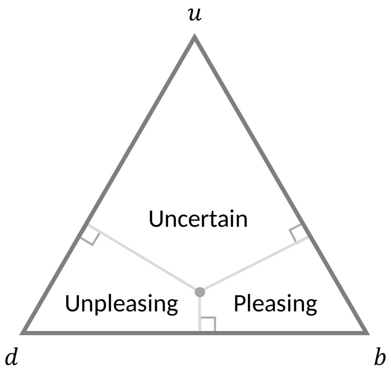


FIGURE 6. **Class boundaries of the ternary classification on the equilateral triangle of b, d, and u. The center point is set to the median value of each component for the training dataset, which is $(b, d, u) = (0.429, 0.457, 0.119)$.**

Consider an application where aesthetic binary classification of images is performed to predict whether each image will be aesthetically "pleasing" or "unpleasing". Even if an image is predicted to be aesthetically pleasing in terms of the mean rating, if a nonnegligible proportion of users would disagree on the prediction, then it would be better not to present the image as pleasing to users in order to maximize the users' satisfaction. In other words, using ternary classification that includes the third class "uncertain" can increase user satisfaction.

We perform ternary classification to classify an image as pleasing, unpleasing, or uncertain, based on the three probabilities obtained from our subjectivity modeling framework. The boundaries between the three classes are defined on the equilateral triangle of the three probabilities as shown in Fig. 6. The center point where the boundaries meet is set by the median values of b, d, and u for the training dataset. One may consider to simply set it to $(b, d, u) = (1/3, 1/3, 1/3)$, but we found that the number of images in each class becomes significantly unbalanced. Thus, we use the median values instead.

We simulate the two rules, i.e., the conventional binary classification and our ternary classification, on the test data of the AVA dataset using the trained ConvNeXT-T model. Each of the ratings is considered as a rater. The performance of the two rules is measured by the satisfaction ratio, which is calculated as the average proportion of raters whose ratings (pleasing or unpleasing) are the same to the prediction result. The obtained satisfaction ratios are 63.70% and 65.52% for the baseline rule and our rule, respectively. This improvement demonstrates the effectiveness of our framework in this application.

VII. CONCLUSION

We proposed a probabilistic framework for modeling image aesthetics, particularly considering proper quantification of the subjectivity. The framework allowed us to model the rating distribution as a beta distribution and to obtain the probabilities of being definitely pleasing, being definitely unpleasing, and being uncertain simultaneously. Through the framework, the image aesthetic prediction performances are improved. The probability of being uncertain was used to define AesU, which is an intuitive subjectivity metric and also as reliable as STD. Through the experiments, it was shown that AesU is valuable as a subjectivity metric. In the future work, we plan to further explore applications exploiting the uncertainty of image aesthetics.

REFERENCES

[1] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2408–2415.

[2] C. Kang, G. Valenzise, and F. Dufaux, "Predicting subjectivity in image aesthetics assessment," in *Proc. IEEE 21st Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2019, pp. 1–6.

[3] A. Jøsang, *Subjective Logic*, vol. 3. Berlin, Germany: Springer, 2016.

[4] C. Li and T. Chen, "Aesthetic visual quality assessment of paintings," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 236–252, Apr. 2009.

[5] E. Mavridaki and V. Mezaris, "A comprehensive aesthetic quality assessment method for natural images using basic rules of photography," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 887–891.

[6] L. Sun, T. Yamasaki, and K. Aizawa, "Photo aesthetic quality estimation using visual complexity features," *Multimedia Tools Appl.*, vol. 77, no. 5, pp. 5189–5213, Mar. 2018.

[7] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "RAPID: Rating pictorial aesthetics using deep learning," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 457–466.

[8] L. Mai, H. Jin, and F. Liu, "Composition-preserving deep photo aesthetics assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 497–506.

[9] D. Liu, R. Puri, N. Kamath, and S. Bhattacharya, "Modeling image composition for visual aesthetic assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 320–322.

[10] K. Sheng, W. Dong, M. Chai, G. Wang, P. Zhou, F. Huang, B.-G. Hu, R. Ji, and C. Ma, "Revisiting image aesthetic assessment via self-supervised feature learning," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 5709–5716.

[11] H. Zeng, Z. Cao, L. Zhang, and A. C. Bovik, "A unified probabilistic formulation of image aesthetic assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 1548–1561, 2020.

[12] J.-T. Lee and C.-S. Kim, "Image aesthetic assessment based on pairwise comparison—A unified approach to score regression, binary classification, and personalization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1191–1200.

[13] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "MUSIQ: Multi-scale image quality transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5128–5137.

[14] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.

[15] Q. Chen, W. Zhang, N. Zhou, P. Lei, Y. Xu, Y. Zheng, and J. Fan, "Adaptive fractional dilated convolution network for image aesthetics assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14102–14111.

[16] J. H. Ching, J. See, and L.-K. Wong, "Learning image aesthetics by learning inpainting," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 2246–2250.

[17] D. She, Y.-K. Lai, G. Yi, and K. Xu, "Hierarchical layout-aware graph convolutional network for unified aesthetics assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8471–8480.

[18] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "MaxViT: Multi-axis vision transformer," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 459–479.

[19] Y. Ke, Y. Wang, K. Wang, F. Qin, J. Guo, and S. Yang, "Image aesthetics assessment using composite features from transformer and CNN," *Multimedia Syst.*, vol. 29, no. 5, pp. 2483–2494, Oct. 2023.

[20] L. Li, T. Zhi, G. Shi, Y. Yang, L. Xu, Y. Li, and Y. Guo, "Anchor-based knowledge embedding for image aesthetics assessment," *Neurocomputing*, vol. 539, Jun. 2023, Art. no. 126197.

[21] T.-S. Park and B.-T. Zhang, "Consensus analysis and modeling of visual aesthetic perception," *IEEE Trans. Affect. Comput.*, vol. 6, no. 3, pp. 272–285, Jul. 2015.

[22] W.-H. Kim, J.-H. Choi, and J.-S. Lee, "Objectivity and subjectivity in aesthetic quality assessment of digital photographs," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 493–506, Jul. 2020.

[23] J. Ren, X. Shen, Z. Lin, R. Mech, and D. J. Foran, "Personalized image aesthetics," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 638–647.

[24] G. Wang, J. Yan, and Z. Qin, "Collaborative and attentive learning for personalized image aesthetic assessment," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 957–963.

[25] L. Li, H. Zhu, S. Zhao, G. Ding, H. Jiang, and A. Tan, "Personality driven multi-task learning for image aesthetic assessment," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2019, pp. 430–435.

[26] H. Zhu, L. Li, J. Wu, S. Zhao, G. Ding, and G. Shi, "Personalized image aesthetics assessment via meta-learning with bilevel gradient optimization," *IEEE Trans. Cybern.*, vol. 52, no. 3, pp. 1798–1811, Mar. 2022.

[27] P. Lv, J. Fan, X. Nie, W. Dong, X. Jiang, B. Zhou, M. Xu, and C. Xu, "User-guided personalized image aesthetic assessment based on deep reinforcement learning," *IEEE Trans. Multimedia*, vol. 25, pp. 736–749, 2023.

[28] Y. Yang, L. Xu, L. Li, N. Qie, Y. Li, P. Zhang, and Y. Guo, "Personalized image aesthetics assessment with rich attributes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 19829–19837.

[29] H. Jang and J.-S. Lee, "Analysis of deep features for image aesthetic assessment," *IEEE Access*, vol. 9, pp. 29850–29861, 2021.

[30] J. A. Hartigan and P. M. Hartigan, "The dip test of unimodality," *Ann. Statist.*, vol. 13, no. 1, pp. 70–84, Mar. 1985.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[33] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976.

HYEONGNAM JANG received the B.S. degree in electrical and electronic engineering from Yonsei University, South Korea, where he is currently pursuing the Ph.D. degree in integrated technology. His research interests include machine learning and image aesthetics.

YEEJIN LEE received the Ph.D. degree in electrical and computer engineering from the University of California at San Diego, La Jolla, CA, USA, in 2017. She is currently an Assistant Professor with the Seoul National University of Science and Technology, Seoul, Republic of Korea. She was a Postdoctoral Fellow in radiology with the University of California at Los Angeles, Los Angeles, CA, USA, from 2017 to 2018. Her research interests include computer vision and machine learning.

JONG-SEOK LEE (Senior Member, IEEE) received the Ph.D. degree in electrical engineering and computer science from the Korea Advanced Institute of Science and Technology (KAIST), Korea. He was a Researcher with the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. He is currently a Professor with the School of Integrated Technology, Yonsei University, South Korea. He has authored or coauthored more than 150 publications. His research interests include multimedia signal processing and machine learning. He serves as an Editor for *IEEE Communications Magazine* and *Signal Processing: Image Communication*.