AN IMAGE AUGMENTATION METHOD FOR QUALITY ASSESSMENT DATABASE

[†]*Yucheng Zhu,* [†]*Guangtao Zhai,* [†]*Wenhan Zhu and* [‡]*Jiantao Zhou*

[†]Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong University, Shanghai, China
[‡]Department of Computer and Information Science, University of Macau, Macau
{zyc420, zhaiguangtao ,zhuwenhan823}@sjtu.edu.cn, jtzhou@umac.mo

ABSTRACT

Image databases for quality assessment are helpful to evaluate the performance of objective assessment methods. Recommendations in regard to the constitution of databases and experimental methods of the subjective assessment have been proposed to ensure the database a good ground truth for the validation of objective quality assessment methods. However, these restrictions make databases scale-limited by covering small number of scenes distorted by few levels. To enrich IQA databases and increase the generalization capability of IQA models, we devise an effective image augmentation method. The two-stages scheme consists of the image-label pairs generation by minimizing the free energy between the pristine image and its augmentation as well as the distortion level interpolation which is based on the monotonicity of the perceptual quality with the severity of distortion. The experimental results show the ability of the augmented database to improve the prediction accuracy of learning-based no-reference image quality assessment metrics which in turn demonstrates the effectiveness of our method.

***Index Terms*—** Image quality assessment (IQA), no-reference, image augmentation, generalization capability

I. INTRODUCTION

With the rapid development of visual acquisition and display technologies, digital images are ubiquitous in our life. Those images often suffer from quality degradation caused by various distortions, hence the devising of effective image quality assessment (IQA) algorithms remains an important topic of research. In the prevalent effective no-reference (NR) IQA metrics, it is a common strategy to employ the natural scene statistics (NSS) [1], [2], [3], [4], [5], [6]. And some of these models that employ statistical features to measure the severity of distortions follow the two-stages scheme: extraction of statistical features, then regression model learning. It has been observed that after suitable normalization, NSS based features follow statistical laws well.

However, the small scale of current IQA databases causes some problems, among which the unsuitability to serve as the training dataset for objective assessment metrics is common. In IQA databases, the distorted images are generated by adding different levels of distortions into reference images. The subjective measurements will provide the mean opinion scores (MOSs) or the difference of quality scores (DMOSs) which can be used as the ground truth labels. To obtain the accurate labels, the formations of IQA databases should be under the guidelines of recommendations regarding the subjective assessment methods [7], [8], [9]. The images should be evaluated in one session to minimize the scale mismatches. The time of subjective assessment should be designed to minimize the effect of observer fatigue. And the introduced distortion levels should guarantee enough perceptual difference for the accurate rank order. These factors can ensure the database a good ground truth for the validation of objective quality assessment methods but also make the database to be scale-limited.

More detailedly, the number of undistorted images in databases is far from enough to cover natural scenes, and the number of distortion levels for pristine images is also limited [10]. The small number of scenes and the sparsity of distortion levels are liable to generate scene-specific or distortion level-specific features, and the employment of these features in objective IQA methods is liable to decrease their generalization capability. For learning based methods, these problems are likely to cause overfitting during the model training which tends to happen when the amount of training data is not large enough to support model's expressive power. These are inevitable problems of the recommendations regarding the subjective assessment methods. In this paper, we devise an effective image augmentation method to help enrich the existing IQA databases. The method consists of two main parts which are the generation of augmentations from pristine images and the interpolation of distortion levels. We resort to the free-energy principle [11] to obtain the perceptually similar images as the augmentation of the source image by minimizing the free energy between the pristine image and its augmentation. The interpolation of distortion level is based on the monotonicity of the perceptual quality with the severity of distortion [12]. The shape of the relation curve is empirically determined by the properties of HVS. And the effectiveness of augmented

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

image-label pairs is demonstrated by their ability to improve the prediction accuracy of learning based methods.

The rest of this paper is organized as follows. Section II first presents the proposed database augmentation method. In Section III, the effectiveness of the method is proved by comparisons of the experimental results. Finally, concluding remarks are given in Section IV.

II. IQA DATABASE AUGMENTATION METHOD

Given an image I as the visual stimulus, the free energy principle suggests an internal generative model $\mathcal{G}$ to govern the cognitive process in the brain. The model $\mathcal{G}$ can adapt itself to different scenes by varying the parameter vector θ [13]. The perception of image I by the brain can be simulated by the integration defined in Eq. (1).

$$-\log P(I|\mathcal{G}) \quad = \quad -\log \int P(I,\theta|\mathcal{G})d\theta. \tag{1}$$

The brain's working can be represented by the internal generative model $\mathcal{G}$ and the behavior of the model can be characterized by parameter θ. For the purpose of calculation, we choose mathematical model $G^{'}$ to simulate the internal generative model and let $Q(\theta|I) = P(\theta|I, \mathcal{G}^{'})$. The latent assumption $\mathcal{G}$ of model can be dropped for simplicity. According to Jensen's inequality we have

$$-\log P(I) \quad \leq \quad -\int Q(\theta|I)\log\frac{P(I,\theta)}{Q(\theta|I)}d\theta. \tag{2}$$

Then we define the right side of Eq. (2) as the free energy:

$$F(I,\theta) \quad = \quad -\int Q(\theta|I)\log\frac{P(I,\theta)}{Q(\theta|I)}d\theta. \tag{3}$$

By the chain rule, $P(I, \theta) = P(I|\theta)P(\theta)$ and we can write (3) into

$$\begin{aligned} F(I,\theta) &= \int Q(\theta|I)\log\frac{Q(\theta|I)}{P(I|\theta)P(\theta)}d\theta \\ &= -\int Q(\theta|I)\log P(I|\theta)d\theta + \int Q(\theta|I)\log\frac{Q(\theta|I)}{P(\theta)}d\theta \\ &= E_Q[\log P(I|\theta)] + KL\left(Q(\theta|I)\|P(\theta)\right). \end{aligned} \tag{4}$$

The KullbackLeibler divergence is a measurement between the approximated densities and the true prior of the model parameters. The term $E_Q[\log P(I|\theta)]$ is the weight-averaged likehood over the approximated posterior density. And the θ can be derived by $\hat{\theta} = \arg\min F(\theta|I)$.

Here we hypothesize the 2D linear autoregressive (AR) model as the mathematical expression of the internal generative model $\mathcal{G}$ for the high description capability of AR model for natural images [13]. The AR model is defined as

$$x_n \quad = \quad \boldsymbol{\chi}^k(x_n)\boldsymbol{\alpha}(x_n) + \varepsilon_n \tag{5}$$

where $\boldsymbol{\chi}^k(x_n)$ is a row-vector that is formed by k neighbors of the nth pixel x_n, $\boldsymbol{\alpha}(x_n) = (\alpha_1, \alpha_2, \alpha_3, ..., \alpha_k)^T$ is the coefficients of AR model and the ε_n is the residual. When the assumed mathematical model well conforms to the internal generative model and when the sample is large enough, from Eq. (4) we can estimate the free energy as the total description length of image I [14]. So we can estimate the model coefficients by minimizing the description length [15].

$$\hat{\boldsymbol{\alpha}} \quad = \quad \arg\min_{\alpha}(-\log P(I|\boldsymbol{\alpha}) + \frac{k}{2}\log N) \tag{6}$$

where N is the data sample size. We fix the order of the model and thus coefficients of AR model can be estimated by residual minimization

$$\hat{\boldsymbol{\alpha}} \quad = \quad \arg\min_{\alpha} ||\mathbf{x_s} - \mathbf{X}\boldsymbol{\alpha}||_2 \tag{7}$$

where $\mathbf{x_s} = (x_1, x_2, ..., x_N)^T$ and $\mathbf{X}(n) = \boldsymbol{\chi}^k(x_n)$. The AR model coefficients can be estimated as $\hat{\boldsymbol{\alpha}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T x$. In this case, the free energy can be measured by the entropy of the residual and we employ the free energy as the supervisor of the pristine image augmentation. The parameters of the label-preserving transformations is best defined by minimizing the free energy between the image and its transformation.

$$\hat{\eta} = \arg\min_{\eta}[F(T_\eta(I)) - F(I)] \tag{8}$$

where T is the transformation and η is the parameter set of T. We employ three transformations to create extra new image-label pairs from pristine images: the image rotation, the image scaling and the image shifting. The parameters η are determined by Eq. (8) to ensure these processes to be label-preserving. In this paper, the rotation angle, the number of pixels to be shifted and the standard deviation of gaussian kernal used in the low-pass filter before image scaling down are ascertained by restricting the difference of free energy within 5%. Under this circumstance, the attached labels of the transformed images are assumed to be the same with the originals.

After the generation of extra new image-label pairs, we interpolate the levels of three common distortions caused by Gaussian noise, Gaussian blur and JPEG compression.

$$P(z) = \frac{1}{\sigma_n\sqrt{2\pi}}\exp(-\frac{(z-\mu)^2}{2\sigma_n^2}) \tag{9}$$

$$G(x,y) = \frac{1}{2\pi\sigma_b^2}\exp(-\frac{x^2+y^2}{2\sigma_b^2}) \tag{10}$$

The probability density function of Gaussian noise term is given by Eq. (9). We control the distortion level by changing the value of standard deviation σ_n. The blurred images are filtered using a 2-D Gaussian kernel given by Eq. (10) of standard deviation σ_b which controls the distortion level. In JPEG compression scheme [16], the quality factor Q can be regarded as a proxy of visual appearance of the final output, the scaling factor S is computed as

$$S = (Q < 50)?\left(\frac{5000}{Q}\right) : 200 - 2Q \tag{11}$$

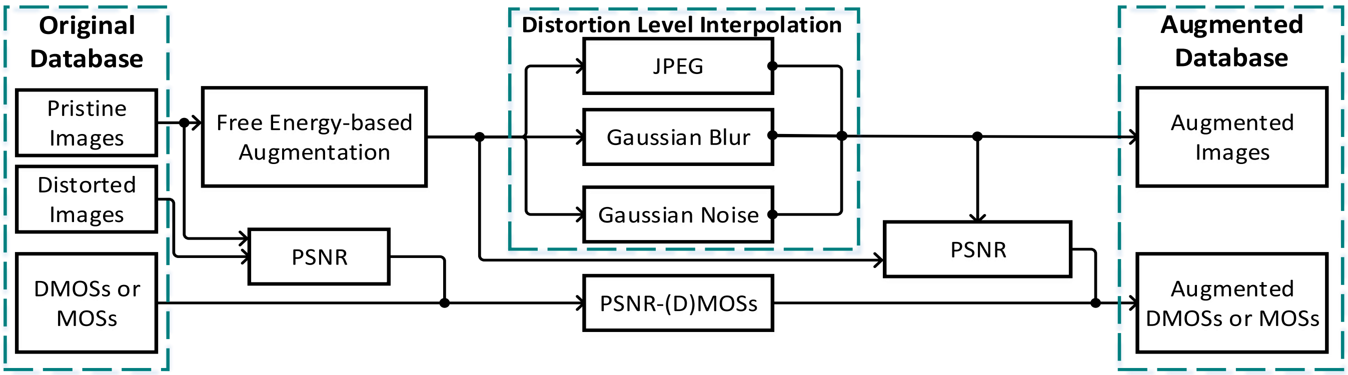

Fig. 1: The diagram of the proposed IQA database augmentation method: Three common distortions including the White noise, Gaussian blur and JPEG compression are analysed. New image-label pairs are generated from the original database by the free energy-based augmentation and distortion level interpolation to enrich the database.

and the scaled quantization table T_s is computed from the original table T_o as

$$T_s = \left\lfloor \frac{S * T_o + 50}{100} \right\rfloor \quad (12)$$

where the T_s alters with the quality factor. The distortion levels are interpolated by varying the parameters of distortion generation functions. By now the images in the augmented database are generated.

However, the corresponding labels are not determined. To obtain corresponding DMOSs (here we suppose that DMOS is the label), we calculate the PSNR of images from the augmented database and get the $(PSNR, Parameter)$ datasets for three kinds of distortions.

$$U(k) = \{(P_i^k, Pa_i^k), i = 1...M\}, k = 1, 2, 3 \quad (13)$$

where M is the number of overall images in augmented database. To map the parameters in Eq. (13) to DMOSs, we need to obtain the relation between PSNR and DMOSs. We can choose any one kind of distortion type for analysis. By calculating the PSNR of images from the original database, we obtain the $(PSNR, DMOS)$ dataset for original database,

$$V = \{(P_i, D_i), i = 1...N\}. \quad (14)$$

To obtain the nonlinear regression function between PSNR and DMOSs, we employ the kernel-based regression defined as:

$$D^* = \frac{\sum_{i=1}^{N} K(P^*, P_i) D_i}{\sum_{i=1}^{N} K(P^*, P_i)} \quad (15)$$

$$K(P^*, P_i) = \exp\left(-\frac{(P^* - P_i)^2}{2h^2}\right) \quad (16)$$

where (P^*, D^*) is the point to be predicted and (P, D) is the point from the dataset in Eq. (14). Due to the training data is from the whole dataset, the regression function characterizes the global feature. To obtain the specific regression functions of different scenes, we combine some data pairs sampled from the global curve with data pairs of each scene in Eq. (14) to form the scene-specific dataset. We also add two apparent points $\{(P \to 0, D = 1), (P \to \infty, D = 0)\}$ into the dataset to modify the regression. The ratio between the number of global points and scene-specific points is empirically set as two to ensure the expression of global characteristics as well as individual characteristics.

$$S(j) = \{(P_i^j, D_i^j), i = 1...2 + r_1 + r_2\}, j = 1...C. \quad (17)$$

where C is the number of scenes in Eq. (14), r_1 and r_2 are the number of global and scene-specific points. By applying the regression defined by Eq. (15) and (16) to the datasets in Eq (17), we can obtain the scene-specific regression function. By now the parameter of each image can map to the DMOS. In other words, we attach each augmented image a DMOS as the label. The label can be easily changed to MOS according to the original database. It is worth to mention that during regressions the h in Eq. (16) should be properly set to ensure the monotonicity of parameter with the DMOS. The diagram of designed augmentation method is shown in Fig. 1.

III. EXPERIMENTAL RESULTS

In this paper, four effective no-reference image quality assessment models including DIIVINE [1], BRISQUE [2], GMLF [3] and NFERM [4] are used to verify the effectiveness of the augmentation method. They are all feature integration based methods including 88, 36, 40 and 23 features respectively. DIIVINE [1] deploys the divisive normalization transform to make the statistics of natural images more Gaussian-like. And features that capture the correlations of coefficients between orientations across scales are extracted. Using these features, they train a distortion-specific model to detect the distortion types and then devise models for each distortion types to evaluate the perceptual quality. BRISQUE [2] employs the mean subtracted contrast normalization strategy to reduce the dependencies.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

Table I: PLCC OF CSIQ-TRAINED MODELS ON LIVE AND TID2013 IMAGE DATABASES.

Database		LIVE BLUR	LIVE NOISE	LIVE JPEG	LIVE ALL	TIDx BLUR	TIDx NOISE	TIDx JPEG	TIDx ALL
DIIVINE [1]	ORIGINAL	0.9053	0.9421	0.8398	0.7387	0.7630	0.8126	0.8180	0.7331
	AUGMENTED	0.9068	0.9560	0.9071	0.8526	0.8089	0.8213	0.8626	0.8011
BRISQUE [2]	ORIGINAL	0.9278	0.9843	0.8970	0.8930	0.8301	0.9012	0.9075	0.8095
	AUGMENTED	**0.9324**	0.9848	0.9324	0.9102	0.8426	0.9157	0.9157	0.8216
GMLF [3]	ORIGINAL	0.9076	0.9756	0.8955	0.8953	**0.8606**	0.8408	**0.9214**	0.8390
	AUGMENTED	0.9230	**0.9895**	0.9277	0.9042	0.8402	**0.9165**	0.8773	**0.8413**
NFERM [4]	ORIGINAL	0.9208	0.9562	0.9008	0.8869	0.8271	0.8740	0.8835	0.8172
	AUGMENTED	0.9251	0.9581	**0.9336**	**0.9152**	0.8251	0.8828	0.9013	0.8212

Table II: SROCC OF CSIQ-TRAINED MODELS ON LIVE AND TID2013 IMAGE DATABASES.

Database		LIVE BLUR	LIVE NOISE	LIVE JPEG	LIVE ALL	TIDx BLUR	TIDx NOISE	TIDx JPEG	TIDx ALL
DIIVINE [1]	ORIGINAL	0.9189	0.9572	0.8442	0.7561	0.7743	0.7986	0.7382	0.7216
	AUGMENTED	**0.9341**	0.9627	0.8944	0.8385	0.8144	0.7999	0.7941	0.7785
BRISQUE [2]	ORIGINAL	0.9326	0.9866	0.8924	0.9098	0.8273	0.8967	0.8608	0.7924
	AUGMENTED	0.9332	**0.9881**	0.9198	**0.9221**	0.8320	0.9078	**0.8968**	0.8154
GMLF [3]	ORIGINAL	0.9336	0.9750	0.8706	0.9096	**0.8414**	0.8362	0.8858	0.8283
	AUGMENTED	0.9223	0.9812	0.9090	0.9157	0.8340	**0.9079**	0.8520	**0.8325**
NFERM [4]	ORIGINAL	0.9276	0.9521	0.8746	0.8921	0.8245	0.8648	0.8621	0.8011
	AUGMENTED	0.9285	0.9594	**0.9241**	0.9095	0.8172	0.8785	0.8902	0.8185

The estimated parameters of the empirical distribution of pairwise products in four orientations across two scales are subsequently used to evaluate the naturalness of images. GMLF [3] proposes joint adaptive normalization operation to make features illumination-unchangeable. They extract features based on the gradient magnitude (GM) and Laplacian of Gaussian (LOG) to predict the image local quality and introduce the dependency index to describe the joint statistics between GM and LOG. NFERM [4] extracts features that can be classified into three groups. The first group consists of features based on free-energy principle and structural degradation model. The second involves HVS inspired features like structural information and gradient magnitude. The third group quantifies the naturalness by fitting the mean subtracted contrast normalized coefficients to generalized Gaussian distribution.

In this paper, three common distortions are chosen for analysis: White noise, Gaussian blur and JPEG compression. The three subsets of CSIQ [17] are used as the training datasets. The CSIQ database contains 30 undistorted images and each kind of distortion involves 5 levels. So there are 150 distorted images for each kind of distortion. We generate new image-DMOS pairs according to the aforementioned augmentation method from the three datasets of CSIQ. The image sets of three distortions in LIVE [18] and TID2013 [19] are used as the testing beds. The LIVE database contains 174, 174 and 233 image-DMOS pairs for White noise, Gaussian blur and JPEG respectively. TID2013 contains 125 pairs for each distortion type. Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SROCC) are used to evaluate performance of our approach. The higher SROCC and PLCC values indicate better performance in terms of correlation with human opinion. PLCC can be considered as a measure of prediction accuracy, while SROCC measures the monotonicity by ignoring the relative distance between the data. The models are separately trained on the original and augmented image sets of each distortion type from CSIQ and are tested on the LIVE and TID2013 databases to validate the effectiveness of augmentation on each distortion type. And overall models that are trained on whole (White noise, Gaussian blur and JPEG) image sets are also compared. With experimental results that are listed in Table 1 and Table 2, we can find the improvement in performance of the model trained on augmented database which in turn demonstrates the effectiveness of our augmentation method for IQA databases.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation of China under Grants 61331014,61521062,61527804, National Hightech R&D Program of China under Grant 2015AA015905 and Macau Science and Technology Development Fund 022/2017/A1.

IV. CONCLUSION

In this paper, we devise an effective image augmentation method. The method consists of two main parts which are the generation of augmentations by minimizing the free energy and by interpolating the distortion levels. The improvement of cross-validation performance on different databases directly demonstrates that our augmented method can increase the generalization capability of learning-based no-reference image quality assessment metrics. By applying our method, the existing database can not only be a good ground truth for the validation of objective quality assessment methods but also be a good database for the design of objective metrics with high generalization capability.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

V. REFERENCES

[1] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, Dec 2011.

[2] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec 2012.

[3] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, Nov 2014.

[4] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 50–63, Jan 2015.

[5] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, "Blind quality assessment based on pseudo reference image," *IEEE Transactions on Multimedia*, 2018.

[6] X. Min, K. Ma, K. Gu, G. Zhai, Z. Wang, and W. Lin, "Unified blind quality assessment of compressed natural, graphic, and screen content images," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5462–5474, Nov. 2017.

[7] "Methodology for the subjective assessment of video quality in multimedia applications," *International Telecommunication Union Recommendation ITU-R BT.1788*, 2007.

[8] "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union Recommendation ITU-R BT.500-13*, 2012.

[9] "General viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays," *International Telecommunication Union Recommendation ITU-R BT.2022*, 2012.

[10] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 1004–1016, Feb 2017.

[11] Karl J Friston, "The free-energy principle: a unified brain theory?," *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[12] G. Zhai and A. Kaup, "Comparative image quality assessment using free energy minimization," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 1884–1888.

[13] Guangtao Zhai, Xiaolin Wu, Xiaokang Yang, Weisi Lin, and Wenjun Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 41–52, 2012.

[14] H. Attias, "A variational bayesian framework for graphical models," *Advances in Neural Information Processing Systems*, vol. 12, pp. 209–215, 2000.

[15] J. Rissanen and G. Langdon, "Universal modeling and coding," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 12–23, Jan 1981.

[16] G. K. Wallace, "The jpeg still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, Feb 1992.

[17] Eric C Larson and Damon M Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 11006, 2010.

[18] H R Sheikh, Wang Z, Cormack L, and A C Bovik, "Live image quality assessment database release 2," [Online]. Available:http://live.ece.utexas.edu/research/quality.

[19] Nikolay N Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir V Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing-image Communication*, vol. 30, pp. 57–77, 2015.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE