Automatic quality prediction of authentically distorted pictures

Deepti Ghadiyaram and Alan Bovik

Biologically inspired computational models automatically predict the quality of any given image, as perceived by a human observer.

Social media and rapid advances in camera and mobile device technology have led to the creation and consumption of a seemingly limitless supply of visual content. However, the vast majority of these digital images are captured by casual amateur photographers whose unsure hands and eyes often introduce annoying artifacts during acquisition. In addition, subsequent storage and transmission of visual media can further degrade their visual quality.

Recent developments in visual modeling have elucidated the impact of visual distortions on perception of such pictures and videos. They have laid the foundation for automatic and accurate metrics that can identify and predict the quality of visual media as perceived by human observers.[1] To address this problem, several objective blind or no-reference (NR) image quality assessment (IQA) algorithms have been developed to predict the perceptual quality of a given (possibly distorted) image without additional information.[2–7] Such quality metrics could be used to monitor and control multimedia services on networks and devices or to prioritize quality of transmission over speed, for example.

Real-world images are usually afflicted by mixtures of distortions that differ significantly from the single, unmixed distortions contained in restrictive and unrepresentative legacy databases.[9–12] We recently designed a unique and challenging image data set with associated human opinion scores called the Laboratory for Image and Video Engineering (LIVE) authentic image quality challenge database[8] (see Figure 1). Using this LIVE challenge database, we have been developing a robust blind IQA model for images suffering from real-world, authentic distortions. We call our model the 'feature maps driven referenceless image quality evaluation engine' (FRIQUEE) index. FRIQUEE outperforms other state-of-the-art blind IQA algorithms on both the LIVE legacy IQA[9] and the LIVE challenge database[8] (see Table 1).
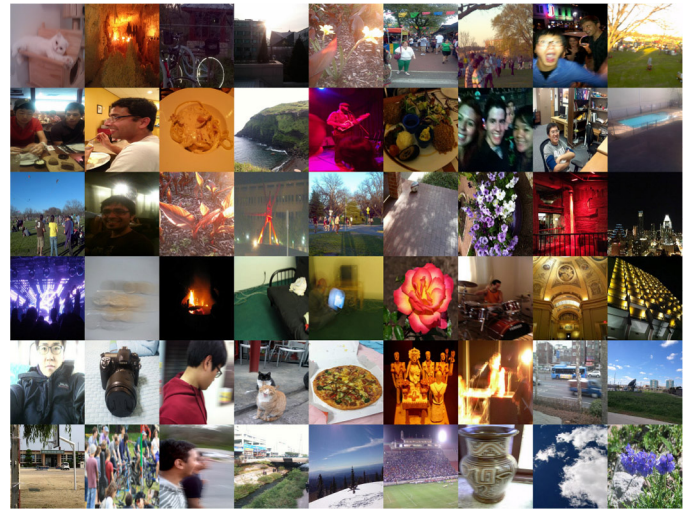

Figure 1. Sample images from the Laboratory for Image and Video Engineering (LIVE) authentic image quality challenge database.[8] This collection comprises 1163 images afflicted with complex mixtures of unknown distortions, of different types and severities, from diverse camera devices, and under varied illumination conditions. The content includes pictures of faces, people, animals, close-up shots, wide-angle shots, nature scenes, man-made objects, images with distinct foreground/background configurations, and images without any notable object of interest.

We have empirically observed that the presence of different mixtures of authentic distortions affects the scene statistics[1] of an image differently from when these distortions occur in isolation.[13] FRIQUEE follows a feature-maps-driven statistical approach, avoiding any latent assumption about the type of distortion(s) contained in an image, focusing instead on the remarkable regularities of the scene statistics of real-world images in the absence of distortions. We capture a large and comprehensive collection of perceptually relevant and reliable 'quality-sensitive' statistical image features from these feature maps that supply us with greater discriminative power on authentic

image distortion ensembles than the most successful NR IQA models. We also leverage the capability of a powerful deep belief network[14] (DBN) to learn a model that maps the image features to complex feature representations that in turn predicts the subjective quality scores more accurately than the state-of-the-art models.

We process every image by first transforming it into color spaces known as RGB, LMS, and CIE LAB, then deriving a set of feature maps by applying operations (including steerable pyramid decomposition in the complex domain,[7] difference of Gaussians,[16] and Laplacian decomposition[17]) on the luminance component and on the four chroma components, i.e., A and B from LAB space, and M and S from the LMS color space (see Figure 2). The definition of each feature map is driven by established perceptual models of the transformations that occur at various stages of visual processing.

We then perform perceptually significant debiasing and divisive normalization operations[15] on each feature map and model the statistical regularities and irregularities exhibited by their histograms using a generalized Gaussian distribution (in real or complex domains)[19] or an asymmetric generalized Gaussian distribution.[20] We compute model parameters, such as shape and variance of these fits, and sample statistics, such as kurtosis and skewness, and use them as image features. A DBN, when
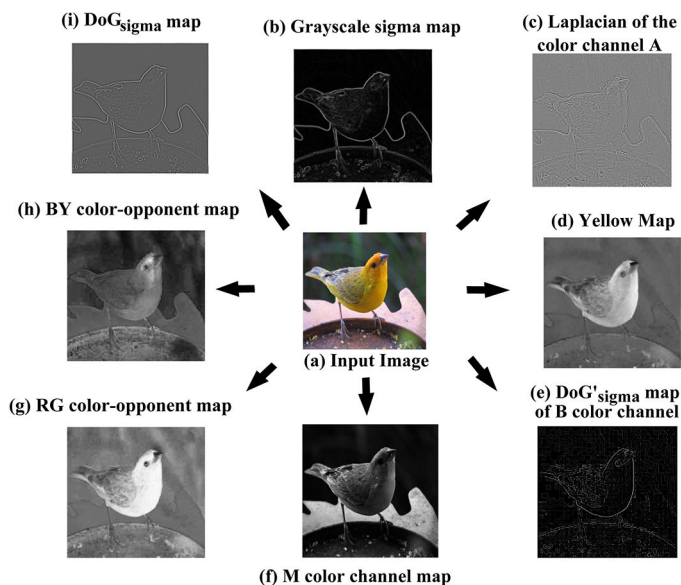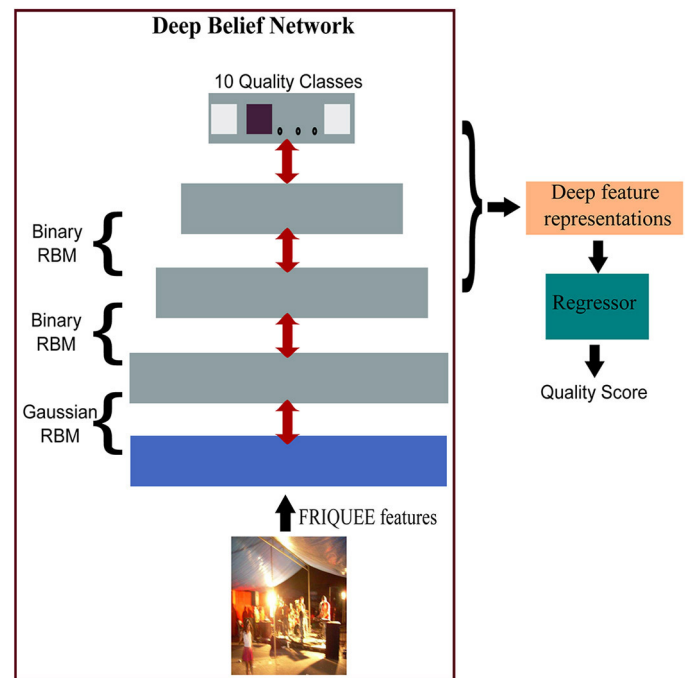

Figure 3. *Configuration of our deep belief network (DBN) model. It has four hidden layers formed by stacking multiple restricted Boltzmann machines (RBMs). Each RBM is trained in a greedy layer-by-layer manner, with the hidden activities of one RBM as the visible input data for training a higher-level RBM. The number of visible units equals the number of features computed on each image. The features from the top three layers serve as 'deep features' to train a support vector regressor (SVR). The unit in the topmost layer of the DBN that is activated determines the quality class.*

Figure 2. *Given (a) any input image, our feature maps driven referenceless image quality evaluation engine (FRIQUEE) first constructs several feature maps in multiple transform domains—some are shown here (b–i)—then extracts scene statistics from these maps after performing perceptually significant divisive normalization[15] on them.*

trained with these image features as input, generates deep feature representations. These deep features are later used to train a support vector regressor (SVR) such that, given a test image's deep features, its quality is predicted (see Figure 3).

We evaluate the performance of several NR IQA algorithms by computing the correlation between the scores predicted by the algorithm and the ground truth subjective scores (see Table 1, which shows that our proposed model combining robust features and a DBN outperforms several other IQA models on unseen test data).

By identifying and addressing the challenges in perceptual quality prediction of images containing mixtures of authentic distortions using FRIQUEE, we have shown that novel techniques push the boundaries of quality prediction powers of state-of-the-art IQA models. To further improve the prediction

Table 1. The median Pearson correlation coefficient (PLCC) and the median Spearman's rank-ordered cross correlation (SROCC) across 50 train-test combinations on the LIVE challenge database.[8] A higher correlation value indicates good performance both in terms of correlation with human opinion as well as the performance of the image quality assessment (IQA) models named in the left column.

	PLCC	SROCC
FRIQUEE	**0.7062**	**0.6824**
BRISQUE [4]	0.6104	0.6018
DIIVINE [3]	0.5577	0.5094
BLIINDS-II [2]	0.4496	0.4049
NIQE [5]	0.4776	0.4210
S3 index [18]	0.3243	0.3054
C-DIIVINE [7]	0.6632	0.6350

power of blind IQA models on real-world distortions, we plan to combine techniques to consider high-level cognitive factors such as semantic content, attention, and aesthetic quality, in addition to accommodating skewness in the real-world distribution of distortions by machine learning.[21–23]

Author Information

Deepti Ghadiyaram and Alan Bovik
The University of Texas at Austin
Austin, TX

Deepti Ghadiyaram is a PhD candidate interested in the intersection of image/video processing and machine learning.

Alan Bovik is the Cockrell Family Regents Endowed Chair Professor and director of LIVE.

References

1. A. C. Bovik, *Automatic prediction of perceptual image and video quality*, **Proc. IEEE 101** (9), pp. 2008–2024, 2013. doi:10.1109/JPROC.2013.2257632
2. M. Saad, A. C. Bovik, and C. Charrier, *Blind image quality assessment: a natural scene statistics approach in the DCT domain*, **IEEE Trans. Image Process. 21** (8), pp. 3339–3352, 2012.
3. A. K. Moorthy and A. C. Bovik, *Blind image quality assessment: from natural scene statistics to perceptual quality*, **IEEE Trans. Image Process. 20** (12), pp. 3350–3364, 2011.
4. A. Mittal, A. K. Moorthy, and A. C. Bovik, *No-reference image quality assessment in the spatial domain*, **IEEE Trans. Image Process. 21** (12), pp. 4695–4708, 2012.
5. A. Mittal, R. Soundararajan, and A. C. Bovik, *Making a 'completely blind' image quality analyzer*, **IEEE Sig. Proc. Lett. 20** (3), pp. 209–212, 2012.
6. H. Tang, N. Joshi, and A. Kapoor, *Learning a blind measure of perceptual image quality*, **IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR)**, pp. 305–312, IEEE, 2011. doi:10.1109/CVPR.2011.5995446
7. Y. Zhang, A. K. Moorthy, D. M. Chandler, and A. C. Bovik, *C-DIIVINE: no-reference image quality assessment based on local magnitude and phase statistics of natural scenes*, **Sig. Proc. Image Commun. 29** (7), pp. 725–747, 2014. doi:10.1016/j.image.2014.05.004
8. D. Ghadiyaram and A. C. Bovik, *Crowdsourced study of subjective image quality*, **Asilomar Conf. Signals Syst. Comput.**, 2014.
9. H. R. Sheikh, M. F. Sabir, and A. C. Bovik, *A statistical evaluation of recent full reference image quality assessment algorithms*, **IEEE Trans. Image Process. 15** (11), pp. 3440–3451, 2006.
10. N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, *TID2008—a database for evaluation of full-reference visual quality assessment metrics*, **Adv. Modern Radio Electron. 10** (4), p. 3045, 2009.
11. E. C. Larson and D. M. Chandler, *Most apparent distortion: full-reference image quality assessment and the role of strategy*, **J. Electron. Imag. 19** (1), 2010.
12. P. Callet and F. Autrusseau, *Subjective quality assessment IRCCyN/IVC database*, 2005. http://www.irccyn.ec-nantes.fr/ivcdb/
13. D. Ghadiyaram and A. C. Bovik, *Perceptual quality prediction on authentically distorted images using natural scene statistics and deep belief nets*, **IEEE Trans. Image Process.**, in preparation, 2015.
14. G. E. Hinton and R. Salakhutdinov, *Reducing the dimensionality of data with neural networks*, **Science 313** (5786), pp. 504–507, 2006. doi:10.1126/science.1127647
15. D. L. Ruderman, *The statistics of natural images*, **Netw. Comput. Neural Syst. 5** (4), pp. 517–548, 1994.
16. H. R. Wilson and J. R. Bergen, *A four mechanism model for threshold spatial vision*, **Vision Res. 19** (1), pp. 19–32, 1979.
17. P. J. Burt and E. H. Adelson, *The Laplacian pyramid as a compact image code*, **IEEE Trans. Commun. 31** (4), pp. 532–540, 1983.
18. C. T. Vu, T. D. Phan, and D. M. Chandler, *S3: a spectral and spatial measure of local perceived sharpness in natural images*, **IEEE Trans. Image Process. 21** (3), pp. 934–945, 2012.
19. K. Sharifi and A. Leon-Garcia, *Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video*, **IEEE Trans. Circuits Syst. Video Technol. 5** (1), pp. 52–56, 1995.
20. N. E. Lasmar, Y. Stitou, and Y. Berthoumieu, *Multiscale skewed heavy tailed model for texture analysis*, pp. 2281–2284, 2009. doi:10.1109/ICIP.2009.5414404
21. L. Bottou and V. Vapnik, *Local learning algorithms*, **Neural Comput. 4** (6), pp. 888–900, 1992. doi:10.1162/neco.1992.4.6.888
22. A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet classification with deep convolutional neural networks*, in F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (eds.), **Adv. Neural Inf. Proc. Syst. 25 (NIPS2012)**, pp. 1097–1105, Curran Associates, 2012.
23. D. Parikh and K. Grauman, *Relative attributes*, **IEEE Int'l Conf. Comput. Vision (ICCV)**, pp. 503–510, 2011. doi:10.1109/ICCV.2011.6126281