# BENCHMARKING STATE-OF-THE-ART VISUAL SALIENCY MODELS FOR IMAGE QUALITY ASSESSMENT

*Wei Zhang[1], Yi Tian[2], Xiaojie Zha[2], and Hantao Liu[1]*

[1]School of Computer Science and Informatics, Cardiff University, Cardiff, UK
[2]International School, Beijing University of Posts and Telecommunications, Beijing, China

## ABSTRACT

A significant current research trend in image quality assessment is to investigate the added value of visual attention aspects. Previous approaches mainly focused on adopting a specific saliency model to improve a specific image quality metric (IQM). It is still not known yet which of the existing saliency models is generally applicable in IQMs; which of the IQMs can profit most/least from the addition of saliency; and how this improvement depends on the saliency model used and the IQM targeted. In this paper, a large-scale benchmark study is conducted to assess the capabilities and limitations of the state-of-the-art saliency models in the context of IQMs. The study provides guidance for the application of saliency models in IQMs, in terms of the effect of saliency model dependency, IQM dependency, and image distortion dependency.

***Index Terms—*** Image quality assessment, visual attention, quality metric, saliency

## 1. INTRODUCTION

Image quality and algorithms for its objective assessment (i.e., image quality metrics (IQMs)) serve as the foundation for optimizing modern digital imaging systems. Being able to automatically and accurately predict image quality as would be perceived by humans continues to be an academic challenge and requires better understanding and modelling of the human visual system (HVS) and its underlying quality evaluation behaviour. To further improve the reliability of IQMs, a significant current research trend is to investigate the impact of visual attention, which refers to a process that enables the HVS to select the most relevant information from a visual scene [1]–[8]. Psychophysical studies have been conducted in an attempt to understand visual attention in relation to image quality assessment [1]–[4]. However, incorporating aspects of visual attention in IQMs remains largely unexplored, while this knowledge would be highly beneficial for further enhancement of IQMs as well as the associated applications.

The vast majority of existing approaches have focused on adding visual attention aspects into IQMs in a rather *ad hoc* way, based on optimizing their performance increase in predicting perceived quality. The concept is generally based on the assumption that distortion occurring in an area that attracts the viewer's attention is more annoying than in any other area; and it intrinsically involves a process in which the local visibility due to distortion is weighted with its corresponding saliency. Various IQMs are extended with the addition of a computational model of visual saliency [5]–[8]. For example, in [5] a saliency model developed in [9] is employed to refine a particular IQM [10] in assessing image quality induced by packet loss. To improve the performance of a sharpness metric [6], a dedicated saliency model is devised and integrated in this metric. There are, however, several essential problems that remain unsolved. First, a variety of saliency models are available in the literature as summarized in [11]–[14]; but the general applicability of these models in the context of IQMs is so far not completely investigated. A rather random selection of a particular saliency model cannot always guarantee an optimized performance gain for a target IQM. Second, it is not known yet whether a saliency model successfully embedded in one specific IQM is also able to enhance the performance of other IQMs, and whether a dedicated combination of a saliency model and an IQM that can improve the assessment of one particular type of image distortion would also improve the assessment of other distortion types. Finally, it has been taken for granted that a saliency model that better predicts human fixations is expected to be more advantageous in improving the performance of IQMs. This speculation, however, has not been statistically validated yet.

In this paper, we draw attention to the need to investigate the concerns raised above, and aim to thoroughly assess the capabilities and potential shortcomings of computational saliency in improving IQM's performance in predicting perceived image quality. By integrating 20 state-of-the-art saliency models into 12 best-known IQMs, we investigate to what extent the amount of performance gain when adding computational saliency to IQMs depends on the saliency model used, IQM targeted, and type of image distortion tested. It also allows us to explore whether or not there is a direct relation between how well a saliency model can predict human fixations and to what extent an IQM can profit from adding this saliency model.

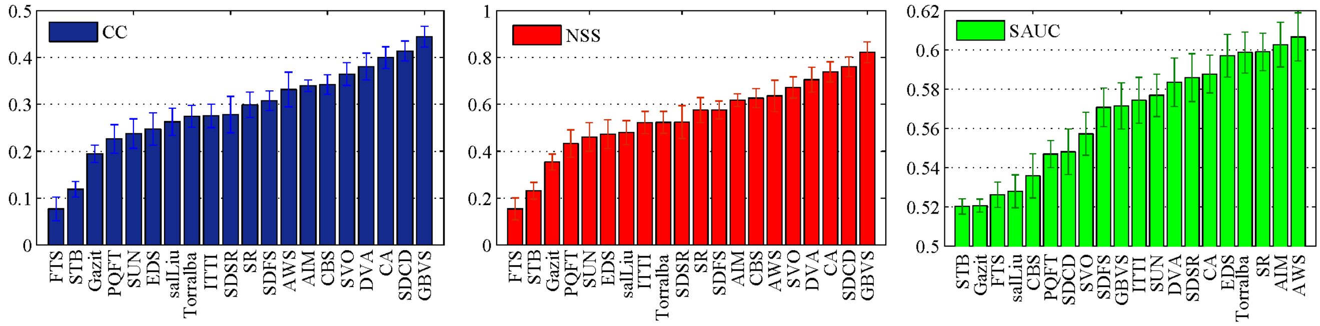


Fig. 1. Illustration of the rankings of visual saliency models in terms of CC, NSS, and SAUC, respectively. CC, NSS, and SAUC are calculated based on the eye-tracking database in [1]. Error bars: 95% confidence interval.

## 2. ADDING COMPUTATIONAL SALIENCY IN IQMS: THE OVERALL EFFECT

### 2.1. Evaluation framework

To assess the added value of computational saliency in IQMs, we follow the general framework established in [15]: the saliency map derived from a saliency model is integrated into an IQM, and the resulting IQM's performance is compared to the performance of the same IQM without saliency. The study is carried out with 20 saliency models and 12 IQMs, which represent the state-of-the-art in the literature.

The IQMs used include eight full-reference (FR) IQMs (i.e., PSNR, UQI, SSIM, MSSIM, VIF, FSIM, IWPSNR and IWSSIM) and four no-reference (NR) IQMs (i.e., GBIM, NPBM, JNBM and NBAM), as detailed in [16]-[18]. Twenty saliency models, namely AIM, AWS, GBVS, SR, DVA, PQFT, SEO, STB, SUN, Torralba, CBS, EDS, FTS, Gazit, CA, ITTI, SDCD, SDFS, salLiu and SVO, as detailed in [11]-[14], are implemented in our study. The evaluation of the performance of IQMs is conducted on the LIVE image quality assessment databases (per image a difference in mean opinion score (DMOS) is derived from an extensive subjective quality assessment study) [19].

To quantify the similarity between a ground-truth human saliency map (HSM) (as described in detail in [1]) obtained from eye-tracking and the modeled saliency map (MSM) derived from a saliency model, three popular measures, namely CC, NSS and SAUC (as described in detail in [11]) are used. The performance of an IQM is quantified by the correlation (i.e., CC and SROCC as prescribed by the video quality experts group [20]) between the outputs of the IQM and the subjective quality ratings (i.e., DMOSs).

### 2.2. The overall effect

The evaluation protocol breaks down into three coherent steps: first, we check the difference in predictability between saliency models; second, by adding these saliency models to individual IQMs we validate whether there is a meaningful gain in performance for the IQMs; finally, we investigate the relation between two trends being the predictability of saliency models and the profitability of including different saliency models in IQMs.

#### *2.2.1. Variation in predictability between saliency models*

The predictability of a saliency model is evaluated against the "ground truth" eye-tracking data in [1], which contains HSMs of 29 original images of the LIVE database. Fig. 1 shows the rankings of saliency models in terms of CC, NSS and SAUC, respectively. It illustrates that the saliency models vary over a wide range of predictability independent of the measure used. Based on SAUC, an analysis of variance (ANOVA) is performed and the results indicate that the numerical differences in predictability between saliency models is statistically significant ($P< 0.01$ at 95% level). Based on this finding, we set out to investigate whether adding these saliency models to IQMs can produce a meaningful gain in their performance, and whether the existence and/or status of such gain is affected by the predictability of a saliency model.

#### *2.2.2. Variation in profitability between saliency models*

Integrating saliency models into IQMs results in a set of new saliency-based IQMs. CC and SROCC are calculated between the subjective DMOS scores and the objective predictions of an IQM. Table 1 summarizes the performance gain (as expressed by the increase in CC, i.e., $\Delta$CC) of a saliency-based IQM over its original version on the LIVE database; and each entry represents the $\Delta$CC averaged over all IQMs. In general this table demonstrates that there is indeed a gain in performance when including saliency models in IQMs. In order to verify whether the effect is statistically significant, hypothesis testing (i.e., Wilcoxon signed rank sum [21]) is conducted using all possible combinations of IQMs and saliency models (i.e., 20 saliency models $\times$ 12 IQMs). The results show that in most cases (i.e., 198 out of 240) the difference in performance between an IQM and its saliency-based derivative is statistically significant. It also tends to suggest that the addition of computational saliency in IQMs makes a meaningful impact on their prediction performance.

Table 1. Performance gain (expressed by the increase in CC, i.e., $\Delta$ CC) between an IQM and its saliency-based version for the images of the LIVE database. Each entry represents the $\Delta$ CC averaged over all IQMs.

| Saliency model | **AIM** | **AWS** | **GBVS** | **SR** | **DVA** | **PQFT** | **SDSR** | **STB** | **SUN** | **Torralba** | **HSM** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Averaged performance gain | 0.012 | 0.008 | 0.021 | 0.025 | 0.018 | 0.021 | 0.022 | -0.009 | 0.015 | 0.015 | **0.02** |
| Saliency model | **CBS** | **EDS** | **FTS** | **Gazit** | **CA** | **ITTI** | **SDCD** | **SDFS** | **salLiu** | **SVO** | |
| Averaged performance gain | 0.009 | 0.016 | 0.006 | 0.01 | 0.021 | 0.007 | 0.021 | 0.014 | 0.013 | 0.011 | |

2.2.3. *Predictability versus profitability*

One could intuitively hypothesize that the better a saliency model can predict human fixations, the more an IQM may profit from adding this saliency model in the prediction of image quality. To check this hypothesis, we calculate the correlation between the predictability of saliency models (based on SAUC in Fig.1) and the averaged performance gain obtained with these models (based on $\Delta$ CC in Table 1) The Pearson correlation is equal to 0.44, suggesting that the relation between the predictability of a saliency model and the actual added value of this model for IQMs is rather weak. Saliency models that are ranked relatively highly in terms of predictability do not necessarily correspond to a larger amount of performance gain when they are added to IQMs. In view of the statistical power, which is grounded on all combinations of 20 saliency models and 12 IQMs, this finding is fairly dependable but indeed surprising; and it suggests that our common belief in the selection of appropriate saliency models for inclusion in IQMs is being challenged. We may conclude that the measure of predictability should not be used as the only criterion to determine the extent to which a specific saliency model can benefit IQMs.

## 3. APPLYING SALIENCY IN IQMS: DEPENDENCIES OF THE PERFORMANCE GAIN

Granted that a meaningful performance gain is in evidence, the actual amount of gain, however, tends to be different for different IQMs, saliency models, and distortion types. Such dependencies of the performance gain have high practical relevance to the application of computational saliency in IQMs, e.g., in circumstances where a trade-off between the increase in performance of an IQM and the expenses needed for saliency modelling is in demand. To this effect, the observed tendencies in the changes of the performance gain are further statistically analyzed in order to comprehend the impact of IQM, saliency model, and the distortion type. The statistical test is based on the original 880 data points of calculated performance gain (i.e., $\Delta$CC in a breakdown version of Table 1, including individual IQMs tested on different distortion types, namely JPEG, WN, GBLUR, JP2K and FF as presented in the LIVE database). A factorial ANOVA is conducted with the performance gain as the dependent variable, and the kind of IQM, saliency model and distortion type as independent variables. The results are summarized in Table 2, and show that all main effects are highly statistical significant.

Table 2. Results of the ANOVA to evaluate the impact of the IQM, saliency model, and image distortion type on the added value of computational saliency in IQMs.

| source | df | F-value | Sig. |
|---|---|---|---|
| saliency model | 19 | 4.036 | .000 |
| distortion | 4 | 32.944 | .000 |
| IQM | 11 | 56.651 | .000 |

Obviously, the kind of IQM has a statistically significant effect on the performance gain. Fig. 2 shows the order of IQMs in terms of the overall performance gain. It illustrates that adding computational saliency results in a marginal gain for IWSSIM, FSIM, VIF, and IWPSNR; the performance gain is either nonexistent or even negative. Compared with such a marginal gain, adding computational saliency to other IQMs, such as UQI, yields a larger amount of performance gain. The difference in the performance gain between IQMs may be attributed to the fact that some IQMs already contain saliency aspects in their metric design but others do not. For example, IWSSIM, VIF, and IWPSNR incorporate the estimate of local information content, which is often applied as a relevant cue in saliency modeling [22].

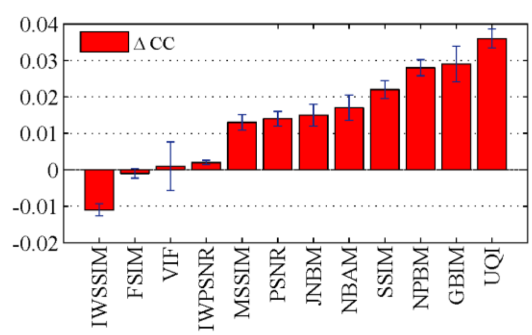


Fig. 2. Illustration of the rankings of IQMs in terms of the overall performance gain (expressed by $\Delta$ CC, averaged over all distortion types and over all saliency models where appropriate) between an IQM and its saliency-based version. Error bars: 95% confidence interval.

There is a significant difference in the performance gain between saliency models. Fig. 3 shows the order of saliency models in terms of the average performance gain obtained by adding individual models to IQMs. A promising gain is found when adding SR, SDSR, PQFT GBVS, and CA to IQMs. The gain achieved for these models is fairly comparable to the gain of adding ground-truth HSM to IQMs. At the other extreme, STB tends to deteriorate the performance of IQMs, and saliency models, such as FTS, do not yield an evident profit for IQMs. STB, which predicts the order in which the eyes move, often highlights the fixation locations (e.g., a certain spot of an object) rather than salient regions (e.g., the entire object). Adding such saliency to IQMs may result in an overestimation of localized distortions. The relatively lower performance gain obtained with FTS is possibly caused by the fact that it segments objects, which are sequentially labeled in a random order. As such, adding saliency in an IQM could randomly give more weight to artifacts in one object than that in another equally salient object.

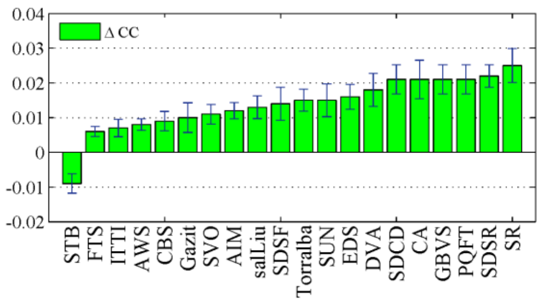


Fig. 3. Illustration of the rankings of the saliency models in terms of the overall performance gain (expressed by $\Delta$ CC, averaged over all distortion types and over all IQMs where appropriate) between an IQM and its saliency-based version. Error bars: 95% confidence interval.

On average, the distortion type has a statistically significant effect on the performance gain, with the order, as shown in Fig. 4. It illustrates that GBLUR profits most from adding computational saliency in IQMs, followed by FF, JPEG, JP2K, and finally WN. Such variation in performance gain may be attributed to the intrinsic differences in perceptual characteristics between individual distortion types. The promising performance gain obtained for GBLUR may be attributed to the fact that adding saliency happens to support the ability of IQMs to distinguish between unintended blur (e.g., on a high-quality foreground object) and intended blur (e.g., in the intentionally blurred background to increase the field of depth).

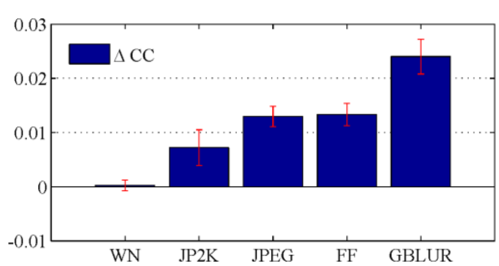


Fig. 4. Illustration of the ranking in terms of the overall performance gain (expressed by $\Delta$ CC, averaged over all IQMs and over all saliency models where appropriate) between an IQM and its saliency based version, when assessing WN, JP2K, JPEG, FF, and GBLUR. Error bars: 95% confidence interval.

## 4. RECIPE FOR SUCCESS

Based on the above-mentioned exhaustive evaluation, guidance on good practices in the application of saliency models in IQMs is provided as follows:

The current soundness of visual saliency modeling is sufficient for IQMs to yield a statistically meaningful gain in their performance. However, the actual amount of performance gain varies among individual combinations of the two variables: saliency models and IQMs.

To decide upon whether a saliency model is in a position to deliver an optimized performance gain for IQMs, we found a threshold value in the overall gain, i.e., 2%, above which the effectiveness of a saliency model, such as SR, SDSR, PQFT, GBVS, CA, and SDCD, is comparable to that of the eye-tracking data and thus is considered to be an optimized amount. Such profit achieved by a saliency model, surprisingly, has no direct relevance to its measured prediction accuracy of human fixations.

When it comes to the issues relating to the IQM dependency of the performance improvement, care should be taken to make a distinction between the IQMs with and without built-in saliency aspects. Adding computational saliency to the former category intrinsically confuses the workings of saliency inclusion, and often produces a smattering of profit. The performance of the latter category of IQMs, however, can be boosted to a large degree with the addition of computational saliency.

The effectiveness of applying saliency-based IQMs in the assessment of different distortion types is subject to the perceptual characteristics of the distortions. The appearance of the perceived artifacts, such as their spatial distribution due to HVS masking, tends to influence the extent to which a certain image may profit from adding saliency to IQMs.

## 5. CONCLUSIONS

In this paper, a benchmark study is conducted to investigate the added value of including computational saliency in objective image quality assessment. Knowledge as the outcome of this paper is highly beneficial for the image quality community to have a better understanding of saliency modeling and inclusion in the context of IQMs. Our findings are valuable to guide developers or users of IQMs to select or decide on appropriate saliency models for their specific application environments.

## 6. REFERENCES


[1] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, Jul. 2011.

[2] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 547–558, Aug. 2010.

[3] U. Engelke, H. Kaprykowsky, H. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 50–59, Nov. 2011.

[4] H. Liu, U. Engelke, J. Wang, P. Le Callet, and I. Heynderickx, "How does image content affect the added value of visual attention in objective image quality assessment?" *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 355–358, Apr. 2013.

[5] X. Feng, T. Liu, D. Yang, and Y. Wang, "Saliency based objective quality assessment of decoded video affected by packet losses," in *Proc. 15th IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 2560–2563.

[6] N. G. Sadaka, L. J. Karam, R. Ferzli, and G. P. Abousleman, "A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling," in *Proc. 15th IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 369–372.

[7] Q. Ma and L. Zhang, "Image quality assessment with visual attention," in *Proc. 15th Int. Conf. Pattern Recognit.*, Tampa, FL, USA, Dec. 2008, pp. 1–4.

[8] R. Barland and A. Saadane, "Blind quality metric using a perceptual importance map for JPEG-20000 compressed images," in *Proc. 13th IEEE Int. Conf. Image Process.*, Atlanta, GA, USA, Oct. 2006, pp. 2941–2944.

[9] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.*, vol. 19, no. 9, pp. 1395–1407, Nov. 2006.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[11] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human model agreement in visual saliency modeling: A comparative study," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 55–69, Jan. 2013.

[12] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Proc. 12th Eur. Conf. Comput. Vis.*, Florence, Italy, Oct. 2012, pp. 414–429.

[13] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. (2014). "Salient object detection: A survey." [Online]. Available: http://arxiv.org/abs/1411.5878.

[14] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. (2015). "Salient object detection: A benchmark." [Online]. Available: http://arxiv.org/abs/1501.02741.

[15] W. Zhang, A. Borji, F. Yang; P. Jiang; H. Liu, "Studying the added value of computational saliency in objective image quality assessment," in *Proc. IEEE Int. Con. VCIP*, Valletta, Malta, Dec. 2014, pp.21-24.

[16] L. Zhang, L. Zhang, X. Mou, D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Proc. 19th Int. Conf. ICIP*, Orlando, FL, USA, Sep. 2012, pp.1477-1480.

[17] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. 9th IEEE Int. Conf. Image Process.*, Rochester, NY, USA, Sep. 2002, pp. III-57–III-60.

[18] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.

[19] H. R. Sheikh, Z. Wang, L. Cormack, and A. Bovik. Live Image Quality Assessment Database Release 2. [Online].Available:http://live.ece.utexas.edu/research/quality , accessed 2006.

[20] Video Quality Experts Group, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II (FR-TV2)," VQEG, Tech. Rep., 2003.

[21] D. C. Montgomery, *Applied Statistics and Probability for Engineers*, 6th ed. New York, NY, USA: Wiley, 2013.

[22] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visualvsearch: An information theoretic approach," *J. Vis.*, vol. 9, no. 3, p. 5, 2009.