https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-120

Interactions Between Saliency and Utility

Edward T. Scott, Sheila S. Hemami; Northeastern University; Boston, Massachussetts, USA

Abstract

We evaluate improvements to image utility assessment algorithms with the inclusion of saliency information, as well as the saliency prediction performance of three saliency models based on successful utility estimators. Fourteen saliency models were incorporated into several utility estimation algorithms, resulting in significantly improved performance in some cases, with RMSE reductions of between 3 and 25%. Algorithms designed for utility estimation benefit less from the addition of saliency information than those originally designed for quality estimation, suggesting that estimators designed to measure utility also measure some degree of saliency information, and that saliency is important for utility estimation. To test this hypothesis, three saliency models are created from NICE and MS-DGU utility estimators by convolving logical maps of image contours with a Gaussian function. The performance of these utility-based models reveals that highly-performing utility estimation algorithms can also predict saliency to an extent, reaching approximately 77% of the prediction performance of state-of-the-art saliency models when evaluated on two common saliency datasets.

Introduction

Image utility estimation is distinctly different from image quality assessment (IQA). While an image's objective quality may be characterized by its similarity to a reference undistorted image, utility is a measure of an image's usefulness to an observer as a proxy for a reference image in the context of accomplishing a task.

Utility is a more useful measure than quality in many applications. For example, firefighters may use thermal imagery to assess risk and devise a plan of action before entering a burning building, and police and military employ both visible light and infrared imagery for surveillance [1]. The objective quality of such images may be low while still allowing observers to accomplish their task [2]. Since image quality is a poor proxy for utility, for these applications measuring utility directly is a better approach [3].

In both the quality and utility estimation paradigms it is common practice to design algorithms consistent with current knowledge of the Human Visual System (HVS). One such technique in the IQA literature is to incorporate visual saliency into existing IQA algorithms, typically by weighting local distortion measures with a modeled saliency map. This approach relies on the assumption that distortion to a visually relevant part of an image impacts perceived quality more significantly than distortion elsewhere. While visual saliency and quality estimation have been studied extensively in relation to the HVS, current understanding of their interaction limits the feasibility of a more comprehensive approach, and this simple methodology can yield small, but significant, performance gains [4, 5, 6].

Given the task-based nature of image utility, saliency seems likely to have greater influence than in the quality realm. In the utility paradigm, distortion to parts of an image which are not relevant to the task a user is completing should have no impact on that image's utility, while in the quality paradigm irrelevant parts of an image have only a reduced impact on quality. Recent work has shown that in addition to object recognition, detection of strong distortion (as seen in the CU-Nantes utility estimation dataset) begins as early as the first set of fixations [7], indicating that there may be a stronger link between fixations and utility than between fixations and quality.

This paper explores the impact of applying saliency weighting to the problem of utility estimation. Fourteen saliency models are incorporated into several utility estimation algorithms and the change in performance over baseline is measured. To assess the degree to which the basic algorithm designs of Natural Image Contour Evaluation (NICE) and Multi-Scale Difference of Gaussian Utility (MS-DGU) account for saliency, these algorithms are modified to generate saliency maps by convolving edge maps (NICE) or logical keypoint maps (MS-DGU) with a Gaussian function having standard deviation similar to the size of the fovea. The performance of these utility-based models is compared to a sampling of saliency models, including the state-of-the-art, on two common saliency datasets developed by Bruce & Tsotsos [8] and Kootstra & Schomaker [9]. For completeness, the three utility-based models are included in the 14 saliency models applied to utility estimation.

Saliency-Weighted Utility Estimation

Five utility estimators were weighted with 14 saliency models to evaluate improvements in utility estimation performance by considering visual saliency. The following sections summarize the utility estimators, saliency models, evaluation approach, and results.

Utility Estimators

The authors have previously developed a utility estimation algorithm called *Multi-Scale Difference of Gaussian Utility (MS-DGU)*, which outperforms other techniques on a dataset containing images labeled with subjective utility ratings collected from human observers [10]. It is based on the assumption that disruption to coarse image structures impairs the HVS's ability to build object representations, thereby reducing image utility. It makes use of the keypoint extraction phase of David Lowe's SIFT algorithm [11], which identifies extrema in a Difference of Gaussian (DoG) decomposition, and evaluates image utility by matching keypoints between test and reference images. Since the publication of [10], the parameters have been fine tuned leading to improved performance.

Prior to MS-DGU, the top-performing utility estimation algorithm was *Natural Image Contour Evaluation (NICE)* [3]. NICE compares contours in a test image with those of the cor-

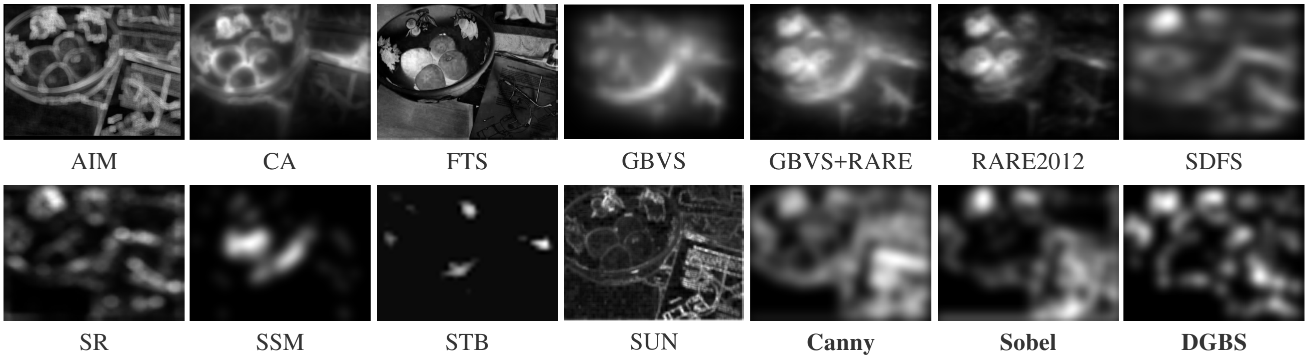


Figure 1. *Example saliency maps. The three maps at bottom right (in bold type face) are generated by models adapted from utility estimators MS-DGU, NICE$_{Canny}$, and NICE$_{Sobel}$, respectively.*

responding reference image, and uses the discrepancy between those contour maps to predict utility. It is designed based on the hypothesis that image utility is directly related to the ability of observers to recognize objects, and perturbing the contours of an image negatively impacts that ability. This is somewhat related to the principles on which MS-DGU is based, but operating on a single scale and using edge detection to identify image contours as opposed to the DoG scale space of MS-DGU (a multi-scale version of NICE was proposed, but did not significantly improve performance). Two versions of NICE are evaluated, with the only difference being the edge detector used (Canny or Sobel).

For comparison, three algorithms originally designed for quality estimation are also evaluated as utility estimators. *PSNR* is a measure of local mean-squared error between test and reference images. It is included here because it is an extremely commonly used, though not particularly well-performing, measure of image quality. The *Structural Similarity Index* (SSIM) is a perception-based measure of degradation of structural information [12]). *Visual Informaton Fidelity (VIF)* is a comparison of an HVS-based statistical measure of image information between a test and reference image [13]. Though VIF was designed as a quality measure, it exhibits similar performance to NICE when used as a utility estimator. VIF and SSIM are the best-performing quality estimation algorithms tested thus far on the utility dataset used in the experiments described below [3].

Saliency Models

Fourteen saliency models representing a sampling of various approaches were applied to the utility estimation algorithms described. They may be summarized as follows:

AIM: An information theoretic approach to saliency modeling with an architecture designed to resemble the visual cortex [8].

CA: Contrast-Aware Saliency Detection considers local low-level features such as contrast and color as well as global information, suppressing common features and promoting unusual features, to identify scene-representative image regions [14].

FTS: Frequency Tuned Salient Region Detection preserves well-defined object boundaries by retaining more frequency content from the original image than other methods [15].

GBVS: Graph-Based Visual Saliency [16] builds activation maps for several feature channels, then normalizes the maps using graph theory.

GBVS+RARE: Results for the mean of GBVS and RARE2012 saliency maps are provided to demonstrate the effect of the Soft Saccadic Model (see below), which takes the GBVS+RARE saliency map as its input.

RARE2012: A multi-scale rarity-based saliency model which identifies the spatial locations in an image with color and orientation features most unlike other parts of the image [17].

SDFS: Saliency Detection Based on Frequency and Spatial Domain Analysis suppresses global regularity in the frequency domain and enhances local features in the spatial domain, then combines these two channels [18].

SR: [19] Saliency Detection: A Spectral Residual Approach computes a saliency map by an inverse Fourier transform of the residual of the log-spectrum of an image after subtracting a generalized average spectrum.

SSM: The Soft Saccadic Model predicts scanpaths and visual fixations by modeling oculomotor biases. It makes these predictions based on a saliency map produced by any saliency model, and also outputs a modified saliency map, though the primary motivation is visual scanpath prediction. The base saliency maps used here are the mean of GBVS and RARE2012 maps as suggested by the authors [20].

STB: The Saliency Toolbox proposes a model based on the Itti Koch algorithm, attempting to infer the locations of proto-objects [21, 22].

SUN: Saliency Using Natural Statistics computes saliency based on self-information of visual features obtained from image statistics collected from a variety of natural images [23].

Canny, Sobel: Saliency maps are generated by convolving a logical edge map produced by the Canny or Sobel edge detector with a Gaussian function. [1]

DGBS: DoG-Based Saliency uses DoG keypoints as a proxy for visual fixations, computing a saliency map by convolving a logical map of keypoints with a Gaussian function. [1]

Example saliency maps produced by each of the algorithms are shown in Figure 1.

Evaluation

The utility and quality estimators described above were evaluated on the CU-Nantes database [24]. This database consists of nine reference 512 × 512 grayscale images and 235 distorted

[1] See "Utility-Based Saliency Models"

versions of the nine originals. Five types of distortion were applied: JPEG compression, DC blocking, JPEG2000+DCQ, Texture Smoothing, and Texture Smoothing + High Pass Filtering. Each of the 235 images in the database is labeled with a subjective utility score derived from a series of paired comparison tests carried out over the course of several experiments with a total of 82 observers. A score below zero indicates an image is not at all useful as a substitute for the reference, and a score over 100 indicates an image is more useful than the corresponding reference.

Objective scores from each estimator were mapped by an affine function to have the same range of values as the collected subjective scores. These objective estimates were then compared to the subjective scores using several metrics, though only Pearson correlation and RMSE are provided here. Performance without saliency weighting is shown in Table 1.

Saliency information was incorporated into the utility estimators using the same approach as [6]. Given an image of size $M \times N$ pixels the distortion measure (DM) of an IQA metric is weighted by a modeled saliency map (MSM), to produce a weighted image quality measure (WIQ):

$$\mathrm{WIQ} = \frac{\sum_{x=1}^{M} \sum_{x=1}^{N} [\mathrm{DM}(x,y) \times \mathrm{MSM}(x,y)]}{\sum_{x=1}^{M} \sum_{x=1}^{N} \mathrm{MSM}(x,y)} \tag{1}$$

The MSM is always generated using the undistorted reference image as input. The DM differs between algorithms. For example, in the case of PSNR, weighting is applied to the local error between test and reference images. For SSIM, the SSIM index map is weighted before averaging. For VIF, weighting is applied in each wavelet subband to all information channels, resizing the MSM to be the same size as each subband [4]. For NICE, weighting is applied to the contour maps. Finally, for MS-DGU, the MSM was decomposed into the DoG domain with the same scale factor as MS-DGU. Each MS-DGU DoG band is weighted by the corresponding MSM DoG band.

Results Suggest New Saliency Models

Tables 2 and 3 summarize the change in utility estimation performance with the incorporation of visual saliency models into the estimators, expressed in terms of Δr (Pearson linear correlation) and $\Delta RMSE$ (as a percentage of baseline). As suggested in [6], the results are tested for statistical significance to ensure that unreliable values are not considered. The errors between estimated utility and ground truth are first tested for normality by measuring their kurtosis. If the kurtosis of the errors is between 2 and 4, they are considered likely to be normally distributed [25].

Table 1: Utility estimation performance on CU-Nantes dataset at baseline (left) and with the inclusion of saliency information (right). r represents Pearson linear correlation. † Quality estimators, but used to estimate utility.

	Baseline		Salience-Weighted	
Estimator	r	RMSE	r	RMSE
PSNR †	0.414	34.08	0.518	32.01
SSIM †	0.843	20.17	0.917	14.96
VIF †	0.943	12.45	0.960	10.43
NICE$_{\text{Sobel}}$	0.924	14.28	0.930	13.74
NICE$_{\text{Canny}}$	0.932	13.61	0.940	12.81
MS-DGU	0.967	9.49	0.969	9.27

The significance of the difference between baseline and saliency weighted errors is determined by a paired sample t-test, in the case of normal error distributions, or a Wilcoxon ranked sum test otherwise. Results with $p < 0.05$ are considered statistically significant, and are indicated by italic font in Table 2 and Table 3. The largest statistically significant improvements are marked in bold, and correspond to column differences in Table 1.

The best significant result for each utility estimation metric is presented in Table 1 alongside baseline performance. While the estimators almost universally improved with the addition of saliency information, results vary widely. However, there is a clear divide between the three quality estimators and the three utility estimators. While saliency-weighting of the quality estimators results in a statistically significant improvement for all but one saliency model, very few saliency models produced a significant result for NICE or MS-DGU. In fact, only eight of the 42 results for the utility estimators are statistically significant, compared to 39 of 42 results for the quality estimators. Clearly, the incorporation of saliency information improves PSNR, SSIM, and VIF as utility estimators, but is not nearly as beneficial for NICE and MS-DGU, which were originally designed for utility estimation. Also of note: NICE$_{\text{Sobel}}$ and NICE$_{\text{Canny}}$ were each most improved by the saliency model based on the other.

One potential explanation for the differing results between quality and utility estimators is that NICE and MS-DGU already measure saliency. If salient image regions are more important to the perception of utility than non-salient regions, an ideal estimator would take that into account. The next section details experiments designed to assess how well NICE and MS-DGU predict saliency.

Utility-Based Saliency Models

MS-DGU is not only the best utility estimator tested, it also benefits the least from incorporating saliency models. Similarly, both forms of NICE were improved marginally by only a few of the many saliency models tested, while PSNR, SSIM, and VIF, all quality estimators, improved significantly. These facts in combination raise the questions: does designing a good utility estimator also entail modeling saliency, and how predictive of saliency are MS-DGU and NICE? Fortunately, the designs of both of these algorithms lend themselves well to adaptation as basic saliency models.

MS-DGU relies on matching keypoints – DoG extrema of sufficient contrast – between test and reference images. Keypoints have high curvature, a trait hypothesized to draw the focus of early stages of the HVS, with surrounding local features forming the basis of object representations [26, 27]. There is therefore reason to suspect that DoG keypoints might be correlated with visual fixations, and by extension predict visual saliency. To identify keypoints, an input image I is decomposed into a DoG scale space with function D:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \tag{2}$$

where G represents a 2D Gaussian function with variance σ^2, $k = 2^{1/3}$, and the decomposition is initialized with $\sigma = 1.6$. Keypoints are identified by comparing each point of $D(x,y,\sigma)$ with its eight neighbors in the current scale, and nine neighbors in the scales above and below. The point is a keypoint if its value is

Table 2: Change in Pearson linear correlation r for a saliency-weighted utility estimator vs. its baseline version when evaluated on the CU-Nantes dataset. † Quality estimators, but used to estimate utility. * Abbreviated for space: GB+RA = GBVS+RARE, RARE = RARE2012. Results with p $<$ 0.05 indicated by italic font. Bold indicates significant result with largest improvement for each estimator. Means are calculated using significant results only.

	AIM	CA	Canny	DGBS	FTS	GBVS	GB+RA *	RARE *	SDFS	Sobel	SR	SSM	STB	SUN	Mean
PSNR †	*0.097*	*0.096*	*0.056*	*0.085*	0.012	*0.074*	*0.091*	***0.105***	*0.068*	*0.103*	*0.100*	*0.095*	*0.010*	*0.074*	*0.081*
VIF †	*0.015*	*0.016*	*0.009*	*0.016*	0.005	*0.014*	*0.016*	*0.017*	*0.013*	*0.015*	***0.017***	*0.015*	*0.004*	*0.012*	*0.014*
SSIM †	*0.070*	*0.070*	*0.056*	*0.063*	-0.016	*0.057*	*0.072*	***0.074***	*0.061*	*0.064*	*0.071*	*0.019*	*0.058*	*0.066*	*0.062*
NICE$_{\text{Sobel}}$	0.004	0.004	***0.006***	*0.005*	*-0.041*	0.001	0.005	0.007	0.001	0.009	0.005	0.008	0.002	0.011	*-.010*
NICE$_{\text{Canny}}$	0.012	0.010	0.017	0.018	*-0.013*	0.009	0.014	0.012	*0.005*	***0.008***	*0.004*	0.012	0.002	0.016	*0.001*
MS-DGU	0.001	*-0.001*	0.001	0.001	0.000	0.002	0.001	0.002	0.002	0.001	-0.001	***0.002***	-0.001	0.000	*0.000*

Table 3: Change in RMSE (in percent of baseline) for a saliency-weighted utility estimator vs. its baseline version when evaluated on the CU-Nantes dataset (negative numbers indicate improvement). † Quality estimators, but used to estimate utility. * Abbreviated for space: GB+RA = GBVS+RARE, RARE = RARE2012. Results with p $<$ 0.05 indicated by italic font. Bold indicates significant result with largest improvement for each estimator. Means are calculated using significant results only.

	AIM	CA	Canny	DGBS	FTS	GBVS	GB+RA *	RARE *	SDFS	Sobel	SR	SSM	STB	SUN	Mean
PSNR †	*-5.59*	*-5.5*	*-3.03*	*-4.8*	-0.59	*-4.14*	*-5.19*	***-6.07***	*-3.72*	*-5.76*	*-5.41*	*-.53*	*-4.1*	*-5.94*	*-4.60*
VIF †	*-13.78*	*-14.71*	*-7.93*	*-14.41*	-4.7	*-12.64*	*-14.82*	*-15.98*	*-11.55*	***-16.26***	*-14.25*	*-3.69*	*-10.44*	*-14.23*	*-12.67*
SSIM †	*-24.01*	*-24.21*	*-18.4*	*-21.42*	4.62	*-18.86*	*-24.71*	***-25.84***	*-20.47*	*-21.77*	*-24.35*	*-5.66*	*-19.35*	*-22.35*	*-20.88*
NICE$_{\text{Sobel}}$	-2.42	-2.61	***-3.85***	*-3.2*	*23*	-0.45	-2.98	-4.36	-0.66	-3.25	-5.31	-1.37	-7.33	-5.89	*5.32*
NICE$_{\text{Canny}}$	-9.07	-7.73	-12.6	-14.16	*8.82*	-6.79	-10.31	-9.18	*-3.85*	*-3.13*	-8.61	-1.11	-12.28	***-5.83***	*-1.00*
MS-DGU	-1.55	*1.46*	-1.44	-1.67	-0.04	-2.52	-1.54	-2.7	-3.38	-2.04	***-2.3***	1.48	-0.35	-1.95	*-0.42*

greater than or less than the values of all 26 neighboring points. Keypoints are thresholded to reject those with low contrast or inconsistent curvature in spatially orthogonal directions, corresponding to suppression of straight edges. MS-DGU uses a contrast threshold of 2.5 for pixel values on the interval $[0,255]$.

DoG-Based Saliency (DGBS) considers all DoG keypoints identified in an image to be a map of visual fixations. To generate an MSM, the fixation map is convolved with a two-dimensional Gaussian function having standard deviation equal to approximately one degree of visual angle, representing an estimate of the size of the fovea [28]. The number of pixels which subtend one degree of visual angle, or pixels per degree (ppd) is dependent on image viewing conditions. While this is a limitation for real-world deployment, these conditions are known for common saliency testing datasets (see Table 4), and the parameter is not overly sensitive.

NICE computes the Hamming distance between dilated logical edge maps of test and reference images. Taking the same approach as DGBS, a saliency map is generated by convolving the logical edge map of an image returned by the Sobel or Canny edge detectors with a Gaussian function of standard deviation equal to one degree. Examples of all three models are shown in Figure 1. The three approaches are generaly more similar than different, all producing relatively "blobby" saliency maps, in contrast with some other approaches which produce more sharply defined image regions.

Evaluation

Following the protocol of other saliency model surveys, to measure similarity between MSMs and human saliency maps (HSM) generated from eye-tracking data, DGBS and the Sobel and Canny edge detector-based models are evaluated using three metrics and an established evaluation platform [6, 29, 30].

Pearson Linear Correlation Coefficient (PLCC) measures the linear correlation between an MSM and the corresponding human saliency map (HSM).

Normalized Scanpath Saliency (NSS) normalizes the MSM to have zero mean and unit standard deviation, then measures the average of normalized MSM values at fixation points of the HSM. $NSS \geq 1$ indicates that modeled saliency is predictive of human fixations, while $NSS \leq 0$ indicates that modeled saliency is random.

Shuffled Area Under Curve (SAUC) is a modification of the classic measurement of area under the receiver operating characteristic curve, where the MSM is considered as a binary classifier to separate a positive set of ground truth human fixation points from a negative set of randomly sampled points. SAUC, instead of uniformly sampling points for the negative set, selects a random subset of human fixations from all other images in the dataset. Human fixations tend to have a roughly Gaussian distribution around the center of an image – as a consequence, saliency model evaluation tends to be center-biased. SAUC is designed to account for this effect and is therefore considered a more rigorous metric than PLCC or NSS [29]. A score of 1 indicates perfect prediction of saliency, whereas a score of 0.5 indicates a model is essentially random.

Each metric is calculated for every image in a dataset, then averaged across all images to generate an overall score for a saliency model on that dataset. Two datasets are considered here: Bruce & Tsotsos and Kootstra & Schomaker. Both are established benchmarks for testing saliency models [8, 9], with statistics as summarized in Table 4. Each dataset contains a similar number of images, but those of Kootsra & Schomaker are higher in resolution and more varied in content – the database contains images from five categories: buildings, nature, flowers, animals, and street scenes. Kootstra & Schomaker is considered a more challenging dataset for this reason.

Results and Discussion

The predictive performance of all 14 saliency models is shown in Figure 2 for all three evaluation metrics on both the Bruce & Tsotsos and Kootstra & Schomaker databases, with the

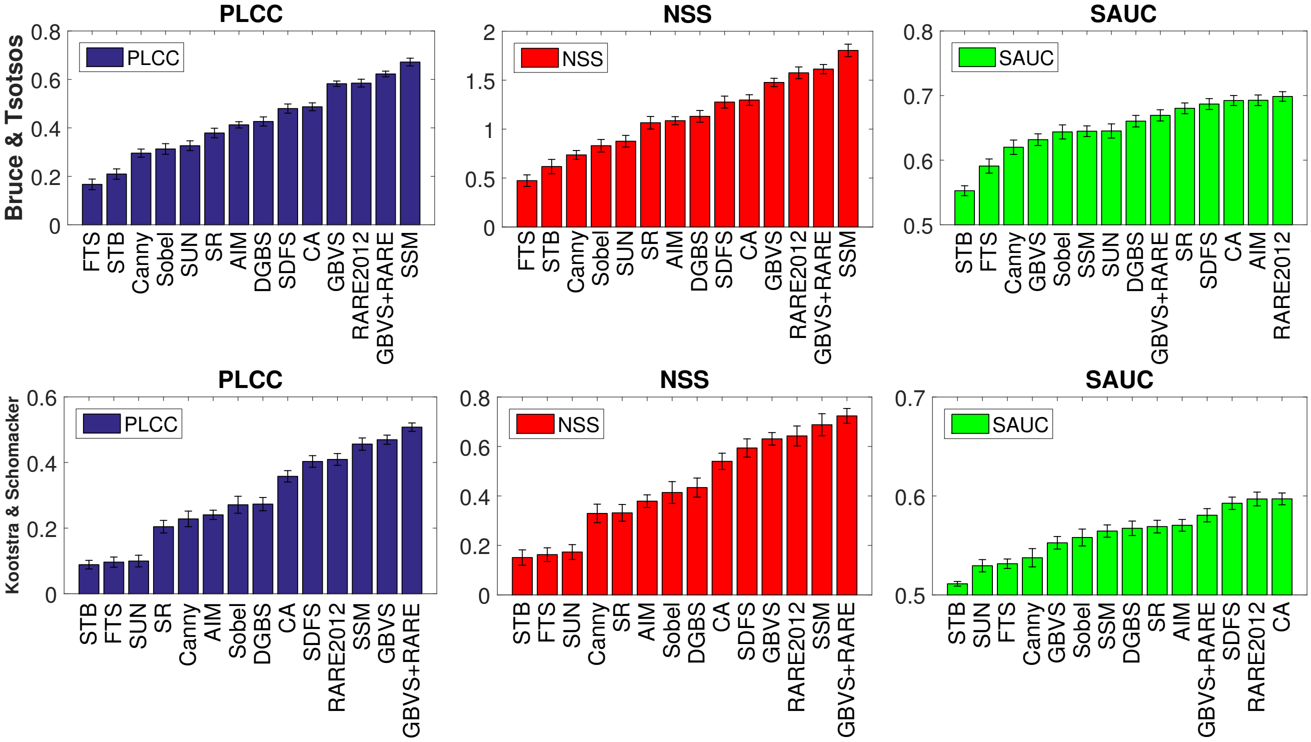


Figure 2. *Performance of visual saliency models on two datasets with three different evaluation metrics: Pearson Linear Correlation Coefficient (PLCC), Normalized Scanpath Saliency (NSS), and Shuffled Area Under Curve (SAUC). Error bars indicate standard deviation of scores across images within each dataset for each saliency model.*

models spanning a wide range of performance. There are significant differences in the ranking of saliency models using the three evaluation metrics. For example, GBVS is ranked significantly lower when evaluated with SAUC. SAUC is designed to normalize center-bias effects to which the other metrics are sensitive, suggesting that GBVS may exhibit some degree of center-bias. None of the utility-based metrics appear to be affected, implying that they are able to predict fixations near image borders as effectively as those near the center. In the case of a utility estimator, that is likely a good thing; objects of interest in a surveillance scenario, for example, are not necessarily in the center region of an image. Due to the lack of sensitivity to center bias, SAUC is considered to be the most rigorous measure of the three, and is the one that will primarily be considered [29].

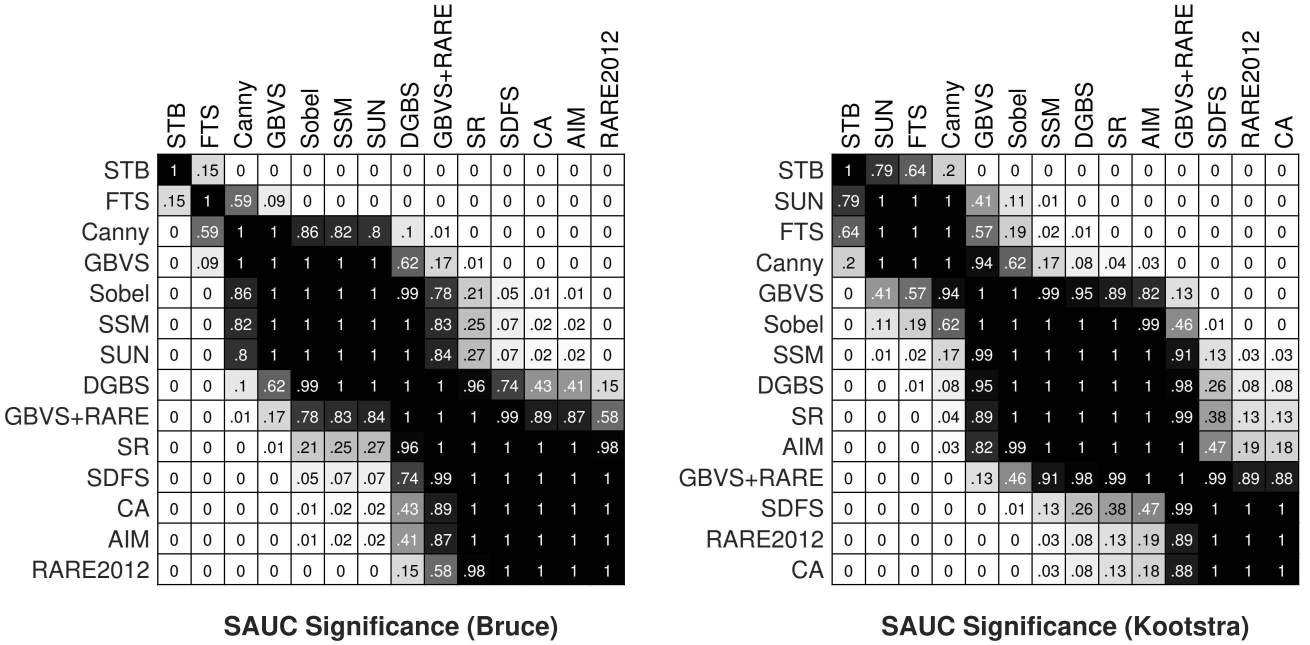
While some models clearly predict utility better than others, there is a lot of overlap and the significance of these differences cannot be determined visually. The kurtosis of PLCC, NSS, and SAUC scores is between 2 and 4 for nearly all saliency models on both datasets, with the exception of NSS scores on the Kootstra dataset, so an analysis of variance (ANOVA) testing methodology may be employed. The kurtosis of NSS values on the Kootstra dataset is high for several models, and ANOVA in that case is unreliable. With each saliency model representing a group, and each group consisting of PLCC, NSS, or SAUC scores for each image in the dataset, the ANOVA strongly indicates the statistical significance of group differences, with F values ranging from 15.6 to 76.1 and p $\approx$ 0. To evaluate the significance of comparisons between individual models, a Tukey's honest significant difference multiple comparison test is conducted between each pair of models. The results are illustrated in Figure 3, which shows p-values for the comparison of SAUC scores for each model pair on both datasets. Similar graphics for PLCC and NSS are not shown, but it should be noted that there is less overlap in terms of significance for those metrics, and there is greater separation between models than for SAUC.

Table 4: Summary of saliency databases used in this evaluation. The acronym *ppd* refers to pixels per degree of visual angle.

Dataset	Images	Observers	Resolution	ppd
Bruce [8]	120	20	681×511	22
Kootstra [9]	100	31	1024×768	34

In general, DGBS is competitive with a wide range of other saliency models, with an SAUC of .6133 when averaged across the Bruce and Kootstra datasets. For comparison, the best performing model, RARE2012, achieves an averaged SAUC of .6479. Taking into consideration the fact that random performance corresponds to an SAUC of 0.5, the saliency prediction performance of DGBS is approximately 77% that of RARE2012, while the Canny and Sobel edge models are less competitive.

Given that average improvements for NICE and MS-DGU were negligible in Table 2, the difference in saliency prediction between the Sobel/Canny and DGBS models may seem surprising. However, considering the different features each technique leverages, it is almost expected. The Canny and Sobel based models are more sensitive to edge content than DGBS, with the



SAUC Significance (Bruce)

	STB	FTS	Canny	GBVS	Sobel	SSM	SUN	DGBS	GBVS+RARE	SR	SDFS	CA	AIM	RARE2012
STB	1	.15	0	0	0	0	0	0	0	0	0	0	0	0
FTS	.15	1	.59	.09	0	0	0	0	0	0	0	0	0	0
Canny	0	.59	1	1	.86	.82	.8	.1	.01	0	0	0	0	0
GBVS	0	.09	1	1	1	1	1	.62	.17	.01	0	0	0	0
Sobel	0	0	.86	1	1	1	1	.99	.78	.21	.05	.01	.01	0
SSM	0	0	.82	1	1	1	1	1	.83	.25	.07	.02	.02	0
SUN	0	0	.8	1	1	1	1	1	.84	.27	.07	.02	.02	0
DGBS	0	0	.1	.62	.99	1	1	1	1	.96	.74	.43	.41	.15
GBVS+RARE	0	0	.01	.17	.78	.83	.84	1	1	1	.99	.89	.87	.58
SR	0	0	0	.01	.21	.25	.27	.96	1	1	1	1	1	.98
SDFS	0	0	0	0	.05	.07	.07	.74	.99	1	1	1	1	1
CA	0	0	0	0	.01	.02	.02	.43	.89	1	1	1	1	1
AIM	0	0	0	0	.01	.02	.02	.41	.87	1	1	1	1	1
RARE2012	0	0	0	0	0	0	0	.15	.58	.98	1	1	1	1

SAUC Significance (Kootstra)

	STB	SUN	FTS	Canny	GBVS	Sobel	SSM	DGBS	SR	AIM	GBVS+RARE	SDFS	RARE2012	CA
STB	1	.79	.64	.2	0	0	0	0	0	0	0	0	0	0
SUN	.79	1	1	1	.41	.11	.01	0	0	0	0	0	0	0
FTS	.64	1	1	1	.57	.19	.02	.01	0	0	0	0	0	0
Canny	.2	1	1	1	.94	.62	.17	.08	.04	.03	0	0	0	0
GBVS	0	.41	.57	.94	1	1	.99	.95	.89	.82	.13	0	0	0
Sobel	0	.11	.19	.62	1	1	1	1	1	.99	.46	.01	0	0
SSM	0	.01	.02	.17	.99	1	1	1	1	1	.91	.13	.03	.03
DGBS	0	0	.01	.08	.95	1	1	1	1	1	.98	.26	.08	.08
SR	0	0	0	.04	.89	1	1	1	1	1	.99	.38	.13	.13
AIM	0	0	0	.03	.82	.99	1	1	1	1	1	.47	.19	.18
GBVS+RARE	0	0	0	0	.13	.46	.91	.98	.99	1	1	.99	.89	.88
SDFS	0	0	0	0	0	.01	.13	.26	.38	.47	.99	1	1	1
RARE2012	0	0	0	0	0	0	.03	.08	.13	.19	.89	1	1	1
CA	0	0	0	0	0	0	.03	.08	.13	.18	.88	1	1	1

Figure 3. *Statistical significance of the difference in SAUC between saliency models evaluated on the Bruce & Tsotsos dataset. Values indicate p-value of comparison between models of each row and column, with $p < 0.05$ indicating significance at the 95% confidence level.*

Canny detector typically returning more edges than Sobel. Particularly in images containing a high amount of detail, this can lead to saliency maps which are too diffuse, while DGBS saliency maps typically include sparser blob-like structures. Images with a detailed subject in front of a smooth background are more common in the Bruce dataset than in the Kootstra dataset, resulting in the relatively poorer performance of the Canny method on Kootstra. The Sobel detector, less sensitive to the busy images of the Kootstra dataset, performs more similarly to DGBS when tested on Kootstra than when tested on Bruce. See Figures 4 and 5 for examples of these differences.

Though in [6] Zhang et. al. reported a weak correlation between the prediction capability of saliency models and average improvement to quality estimators, with a Pearson correlation of 0.44, this experiment yielded different results. The correlation between Δr, averaged over PSNR, SSIM, and VIF (see Table 2) and mean SAUC values of each model is 0.85 (averaged over the two datasets), suggesting that at least in the case of the quality estimators tested, when used to estimate utility, better saliency models are likely to yield greater performance gains.

Conclusion

Incorporating saliency information into quality estimators significantly improved their performance as utility estimators, while the same technique yielded only marginal gains to algorithms designed for utility estimation. Further analysis showed that these utility estimators already measure saliency, potentially accounting for both their insensitivity to saliency weighting and some of the performance differences between unweighted quality and utility estimators. In particular, Difference of Gaussian Based Saliency (DGBS), a saliency model based on the MS-DGU utility estimator, performed comparably to some other recently proposed saliency models, reaching approximately 77% of the performance of state-of-the-art models. While there was relatively strong correlation between saliency prediction performance and Δr for quality estimators, incorporating better performing saliency models into NICE and MS-DGU yielded marginal results. Despite the weaker predictive ability of the DGBS, Canny, and Sobel saliency models, it is possible that they are helped in this case by their tight integration into NICE and MS-DGU. Given the performance of DGBS, further work to refine the use of DoG information for saliency estimation and integrate it into MS-DGU could yield improvements to utility estimation, as well as a more competitive saliency model.

References

[1] Terry L Bisbee and Daniel A Pritchard. Today's thermal imaging systems: background and applications for civilian law enforcement and military force protection. In *Security Technology, 1997. Proceedings. The Institute of Electrical and Electronics Engineers 31st Annual 1997 International Carnahan Conference on*, pages 202–208. IEEE, 1997.

[2] Mikolaj I Leszczuk, Irena Stange, and Carolyn Ford. Determining image quality requirements for recognition tasks in generalized public safety video applications: Definitions, testing, standardization, and current trends. In *Broadband Multimedia Systems and Broadcasting (BMSB), 2011 IEEE International Symposium on*, pages 1–5. IEEE, 2011.

[3] David M Rouse, Sheila S Hemami, Romuald Pépion, and Patrick Le Callet. Estimating the usefulness of distorted natural images using an image contour degradation measure. *JOSA A*, 28(2):157–188, 2011.

[4] Qi Ma and Liming Zhang. Image quality assessment with visual attention. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[5] Xin Feng, Tao Liu, Dan Yang, and Yao Wang. Saliency based objective quality assessment of decoded video affected by packet losses. In *2008 15th IEEE International*

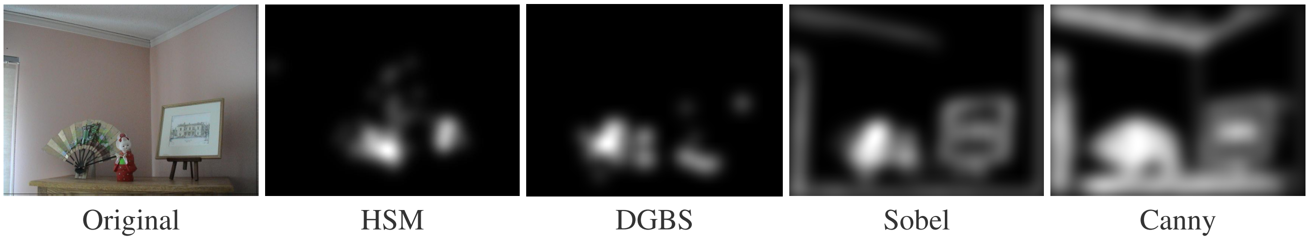


Figure 4. ***Bruce*** *– Original image from Bruce dataset, human saliency map (HSM), and modeled saliency maps using DGBS, Sobel, and Canny methods. The Sobel and Canny methods incorrectly identify salient regions around the framed picture due to the presence of sharp edges.*

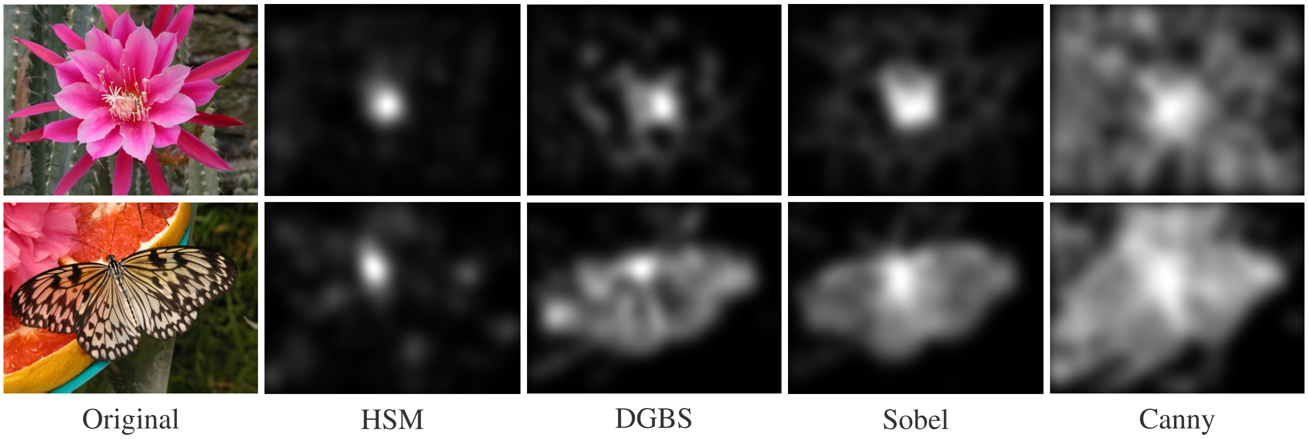


Figure 5. ***Kootstra*** *– Original images from Kootstra dataset, human saliency map (HSM), and modeled saliency maps using DGBS, Sobel, and Canny Methods. DGBS and Sobel produce generally similar results, with DGBS doing a slightly better job of identifying the small, focused regions of the HSM. The saliency maps modeled using the Canny method are much more diffuse due to the Canny edge detector identifying more edges in textured image areas, particularly in the flower image.*

Conference on Image Processing, pages 2560–2563. IEEE, 2008.

[6] Wei Zhang, Ali Borji, Zhou Wang, Patrick Le Callet, and Hantao Liu. The application of visual saliency models in objective image quality assessment: A statistical evaluation. *IEEE transactions on neural networks and learning systems*, 27(6):1266–1278, 2016.

[7] Ernestasia Siahaan, Alan Hanjalic, and Judith A Redi. Does visual quality depend on semantics? a study on the relationship between impairment annoyance and image semantics at early attentive stages. *Electronic Imaging*, 2016(16):1–9, 2016.

[8] Neil DB Bruce and John K Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of vision*, 9(3):5–5, 2009.

[9] Gert Kootstra, Bart de Boer, and Lambert RB Schomaker. Predicting eye fixations on complex visual stimuli using local symmetry. *Cognitive computation*, 3(1):223–240, 2011.

[10] Edward T Scott and Sheila S Hemami. Image utility estimation using difference-of-gaussian scale space. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 101–105. IEEE, 2016.

[11] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[12] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[13] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[14] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal. Context-aware saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):1915–1926, 2012.

[15] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*, pages 1597–1604. IEEE, 2009.

[16] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2006.

[17] Nicolas Riche, Matei Mancas, Matthieu Duvinage, Makiese Mibulumukini, Bernard Gosselin, and Thierry Dutoit. Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, 28(6):642–658, 2013.

[18] Jian Li, Martin Levine, Xiangjing An, and Hangen He. Saliency detection based on frequency and spatial domain analyses. In *Proc. BMVC*, pages 86.1–86.11, 2011. http://dx.doi.org/10.5244/C.25.86.

[19] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE,

2007.

[20] Olivier Le Meur and Zhi Liu. Saccadic model of eye movements for free-viewing condition. *Vision research*, 116:152–164, 2015.

[21] Dirk Walther and Christof Koch. Modeling attention to salient proto-objects. *Neural networks*, 19(9):1395–1407, 2006.

[22] Laurent Itti, Christof Koch, Ernst Niebur, et al. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[23] Lingyun Zhang, Matthew H Tong, Tim K Marks, Honghao Shan, and Garrison W Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of vision*, 8(7):32–32, 2008.

[24] David M Rouse, Romuald Pépion, Sheila S Hemami, and Patrick Le Callet. Image utility assessment and a relationship with image quality assessment. In *IS&T/SPIE Electronic Imaging*, pages 724010–724010. International Society for Optics and Photonics, 2009.

[25] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006.

[26] Daniel Reisfeld, Haim Wolfson, and Yehezkel Yeshurun. Context-free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision*, 14(2):119–130, 1995.

[27] Ilya A Rybak, VI Gusakova, AV Golovan, LN Podladchikova, and NA Shevtsova. A model of attention-guided visual perception and recognition. *Vision research*, 38(15):2387–2400, 1998.

[28] Olivier Le Meur and Thierry Baccino. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods*, 45(1):251–266, 2013.

[29] Ali Borji, Dicky N Sihite, and Laurent Itti. Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Transactions on Image Processing*, 22(1):55–69, 2013.

[30] Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. What do different evaluation metrics tell us about saliency models? *arXiv preprint arXiv:1604.03605*, 2016.

Author Biography

Edward T. Scott received the B.S. and M.S. degrees from Northwestern University in 2010 and worked as an associate technical staff member at MIT Lincoln Laboratory from 2010-2014. He is currently a PhD student at Northeastern University in the department of Electrical and Computer Engineering. His research interests include human visual perception, image analysis, and image quality assessment.

Sheila S. Hemami received the Ph.D. degree from Stanford University (1994) and has held positions at Hewlett-Packard Laboratories, Cornell University, and Northeastern University. Dr. Hemami is a Fellow of the IEEE and has held various leadership positions in the IEEE. She has received numerous college and national teaching awards. Her research interests broadly concern communication of visual information, both from a signal processing perspective and from a psychophysical perspective.