Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: https://orca.cardiff.ac.uk/id/eprint/143106/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Guo, Pengfei, Zhao, Xin, Wu, Dongqing, Zhang, Lei, Zeng, Delu and Liu, Hantao ORCID: https://orcid.org/0000-0003-4544-3481 2021. Convex optimization method for quantifying image quality induced saliency variation. IEEE Access 9 , pp. 111533-111543. 10.1109/ACCESS.2021.3102465 file

Publishers page: https://doi.org/10.1109/ACCESS.2021.3102465 <https://doi.org/10.1109/ACCESS.2021.3102465>

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

information services
gwasanaethau gwybodaeth

# Convex optimization method for quantifying image quality induced saliency variation

**PENGFEI GUO[1,4], XIN ZHAO[2], DONGQING WU[1], LEI ZHANG[3], DELU ZENG[4] AND HANTAO LIU[2]**

[1]School of Computational Science, Zhongkai University of Agriculture and Engineering, Guangzhou, 510115, China
[2]School of Computer Science and Informatics, Cardiff University, Cardiff, CF24 3AA, United Kingdom
[3]College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou, 510115, China
[4]School of Mathematics, South China University of Technology, Guangzhou, China

Corresponding author: Dongqing Wu (dongqing.wu@foxmail.com) and Hantao Liu (LiuH35@cardiff.ac.uk).

**ABSTRACT** Visual saliency plays a significant role in image quality assessment. Image distortions cause shift of saliency from its original places. Being able to measure such distortion-included saliency variation (DSV) contributes towards the optimal use of saliency in automated image quality assessment. In our previous study a benchmark for the measurement of DSV through subjective testing was built. However, exiting saliency similarity measures are unhelpful for the quantification of DSV due to the fact that DSV highly depends on the dispersion degree of a saliency map. In this paper, we propose a novel similarity metric for the measurement of DSV, namely MDSV, based on convex optimization method. The proposed MDSV metric integrates the local saliency similarity measure and the global saliency similarity measure using the function of saliency dispersion as a modulator. We detail the parameter selection of the proposed metric and the interactions of sub-models for the convex optimization strategy. Statistical analyses show that our proposed MDSV outperforms the existing metrics in quantifying the image quality induced saliency variation.

**INDEX TERMS** Image quality assessment, saliency, saliency variation, similarity measure, visual attention

## I. INTRODUCTION

SALIENCY – the scene-driven, bottom-up selective visual attention mechanism of the human visual system – has been widely studied in the area of image quality assessment [1]–[6]. The application of saliency in image quality metrics potentially improves their reliability in predicting image quality as perceived by humans [7]–[14]. However, the optimal use of saliency requires a deeper understanding of how saliency plays a role in image quality assessment and how perception relating to saliency can be effectively quantified.

Eye-tracking studies [15], [16] have been undertaken to better understand the interaction between saliency and image distortions. The SIQ288 database [16] reveals the correspondences between the changes in image quality and the changes in saliency, as an example illustrated in Figure 1. Note all saliency maps contained in the SIQ288 database were rendered from ground truth eye-tracking data. It can be seen from Figure 1 that the difference between the saliency of the "high quality" image and that of the "original image" is marginal; the difference in saliency is modest for the "medium quality" image; whereas the saliency of the "low quality" image significantly differs from that of the "original image". To measure such so-called distortion-induced saliency variation (DSV) – the similarity between the original saliency (e.g., (e) in Figure 1) and the deviated saliency (e.g., (f), (g) or (h) in Figure 1) – a benchmark was established in [17]. In this benchmark, the difference mean saliency variation score (DMSS) per saliency map in the SIQ288 database is yielded via a subjective study. As shown in Figure 1, the higher the DMSS, the less similar the deviated saliency is from the original saliency. The DMSS scores are reliable, however, they are expensive and impractical in many circumstances. A more realistic way to measure saliency

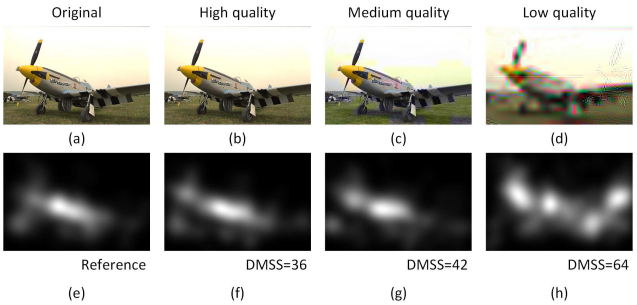



**FIGURE 1.** Illustration of the distortion-induced saliency variation (DSV). DMSS (difference mean saliency variation score) represents the degree of similarity between the deviated saliency map and the reference saliency map.

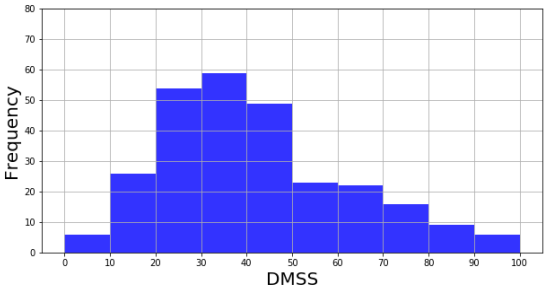



**FIGURE 2.** Illustration of the histogram of the difference mean saliency variation scores (DMSS) of the DSV benchmark [17].

variation (i.e., DSV) is to use computational algorithms.

In the literature, many algorithms have been established and used for similarity measures in various applications. In particular, these similarity metrics are used in the area of saliency modelling to evaluate a computational saliency model's ability to predict human fixations. The plausibility of using these metrics for quantifying DSV, however, is not verified. In this paper, we first analyse the performance of eight state-of-the-art similarity metrics in measuring DSV; and then we propose a new metric for assessing the distortion-induced saliency variation (DSV). The proposed metric is inspired by the characteristics of the human visual system (HVS), using a saliency dispersion measurement as a key component in a convex optimization method to integrate the local saliency similarity and the global saliency similarity.

The rest of this paper is organized as follows. Section II provides an analysis of the state-of-the-art similarity metrics, and the description of the DSV benchmark. Section III details the proposed algorithm. The overall performance of the proposed algorithm is presented in Section IV. Section V concludes the paper.

## II. ANALYSIS OF THE STATE-OF-THE-ART SIMILARITY METRICS

### A. THE DSV BENCHMARK

The benchmark of the distortion-induced saliency variation (DSV) [17] is based on the SIQ288 database [16], which contains images of varying quality (i.e., 18 pristine images and each distorted by 5 distortion types and 3 distortion levels) and their corresponding saliency maps rendered from ground truth human fixations. The DSV benchmark contains 18 reference saliency maps and 270 deviated saliency maps. Sixteen experts in computer vision assessed the deviations of saliency maps of distorted images. The resulting difference mean saliency variation score (DMSS) represents the degree of similarity between each deviated saliency map and its original saliency. The DMSS ranges from 0 to 100, with 0 representing the smallest degree of difference and 100 representing the largest degree of difference. Table 1 lists the details of the distortion-induced saliency variation benchmark. Figure 2 shows the histogram of DMSS scores contained in the benchmark.

**TABLE 1.** Details of the distortion-induced saliency variation (DSV) benchmark [17].

| Characteristics | Information |
|---|---|
| Number of reference saliency maps | 18 |
| Number of deviated saliency maps | 270 |
| Number of different distortion types | 5 |
| Number of different levels for each distortion type | 3 |
| Evaluation method | Simultaneous double stimulus |
| Number of human viewers per image | 16 |

### B. PERFORMANCE OF THE STATE-OF-THE-ART SIMILARITY METRICS

There are many mathematical algorithms that can be used to measure the similarity between two images. In terms of evaluating saliency maps, research has focused on measuring the difference between the predicted saliency (generated by machines) and the ground truth saliency (obtained by eye-tracking) [18]. The popular metrics are: the area under the receiver operating characteristic curve (AUC) [19], [20], Normalized Scanpath Saliency (NSS) [21], Information Gain (IG) [22], SIMilarity (SIM) [23], Pearson's Correlation Coefficient (CC) [24], Kullback-Leibler Divergence (KL) [18], and Earth Mover's Distance (EMD) [25]. It should be noted that each metric has its own advantages and disadvantages depending on the specific applications, for example, some metrics (i.e., AUC, NSS, IG) emphasise on local saliency similarity measurement and some metrics (i.e., SIM, CC, KL, EMD) focus on the global similarity of saliency. However, it is unknow yet whether these metrics are helpful in measuring the distortion-induced saliency variation (DSV). Now, we analyse the performance of these similarity metrics in the DSV context. We calculated the metrics for all test/deviated saliency maps contained in the DSV benchmark. The performance of these metrics in measuring the saliency variation is quantified by comparing the ground truth DMSS scores and the metrics' outputs. Figure 3 shows the scatter plot of the DMSS and each of the similarity metrics. Table 2 lists the quantitative results of the performance in terms of the Pearson's correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), Kendall's rank order

**TABLE 2.** Performance of state-of-the-art similarity metrics for the measurement of distortion-induced saliency variation (DSV) on all saliency maps of the DSV benchmark [17].

| Criterion | AUC-Judd | AUC-Borji | NSS | IG | SIM | CC | KL | EMD |
|---|---|---|---|---|---|---|---|---|
| PLCC | 0.6816 | 0.2787 | 0.6782 | 0.3273 | 0.7475 | 0.7786 | 0.3100 | 0.6121 |
| SROCC | 0.6502 | 0.2106 | 0.6081 | 0.2988 | 0.6962 | 0.7246 | 0.2755 | 0.5502 |
| KROCC | 0.4363 | 0.1373 | 0.4328 | 0.2052 | 0.5063 | 0.5345 | 0.1909 | 0.3832 |

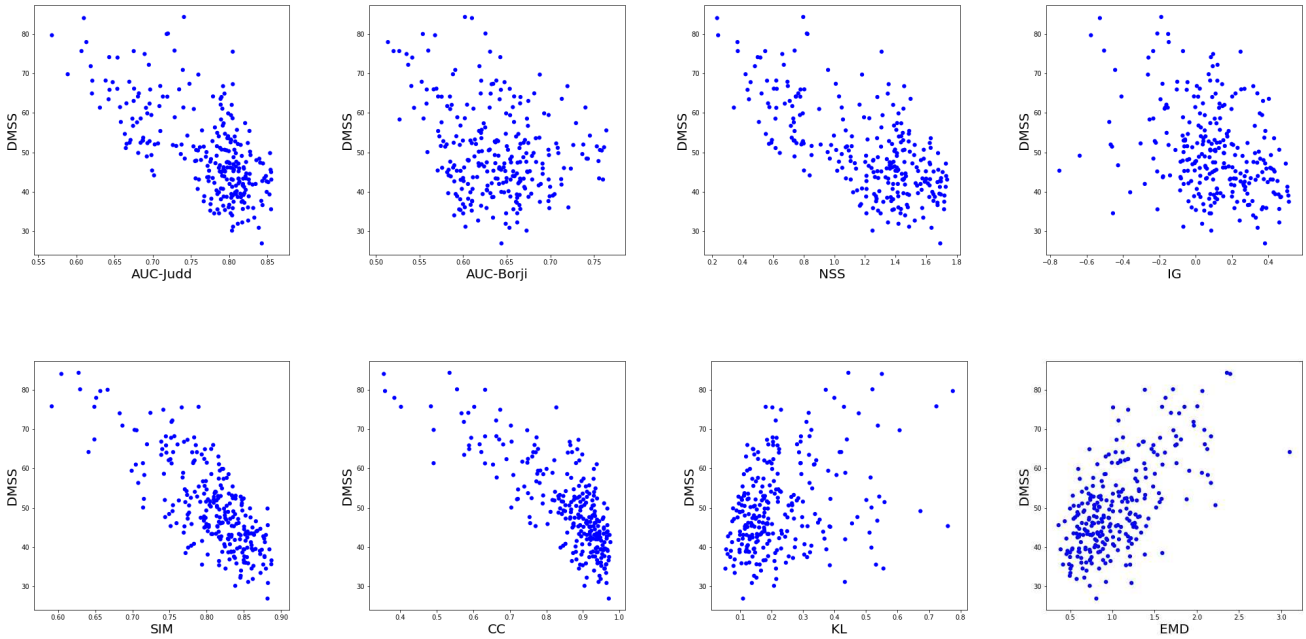



**FIGURE 3.** Scatter plot of the DMSS and each of the similarity metrics (AUC-Judd, AUC-Borji, NSS, IG, SIM, CC, KL and EMD).

correlation coefficient (KROCC), and root mean square error (RMSE) that are formulated as follows:

$$\text{PLCC} = \frac{\sum_{i=1}^{L}(m_i - \bar{m})(n_i - \bar{n})}{\sqrt{\sum_{i=1}^{L}[(m_i - \bar{m})(n_i - \bar{n})]^2}}, \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(m_i - n_i)^2}, \tag{2}$$

$$\text{SROCC} = \frac{\sum_{i=1}^{L}(R(m_i) - \overline{R(m)})(R(n_i) - \overline{R(n)})}{G(m_i, m, n_i, n)}, \tag{3}$$

$$\text{KROCC} = \frac{2[N(\text{concordant pairs}) - N(\text{discordant pairs})]}{L(L-1)}, \tag{4}$$

where $L$ is the number of test saliency maps, $m_i$ and $n_i$ are the metric score and subjective DMSS of the $i$-th saliency map respectively, $\bar{m}$ and $\bar{n}$ are the mean of metric scores and subjective DMSSs respectively, $R(\cdot)$ denotes the rank, $G(m_i, m, n_i, n) = \sqrt{\sum_{i=1}^{L}(R(m_i) - \overline{R(m)})^2}\sqrt{\sum_{i=1}^{L}(R(n_i) - \overline{R(n)})^2}$, $\overline{R(\cdot)}$ is the mean of the ranks, and $N(\cdot)$ represents the amount of corresponding variables.

In general, some metrics (i.e., AUC-Borji, IG, KL) exhibit a poor correlation with the DMSS, meaning they cannot accurately measure the distortion-induced saliency variation. AUC-Judd, NSS, SIM, CC, and EMD metrics show better performance, but there is still room for improvement. Some interesting observations are revealed in our study. For different reference saliency maps from different source images, the responses of the similarity metrics differ. We give two examples here using the correlation matrix visualization. Figure 4 illustrates the correlation matrix of eight similarity metrics and DMSS of all deviated saliency maps from the reference, i.e., "Paint-house". Figure 5 illustrates the correlation matrix for a different reference, e.g., "Man-fishing". The reference saliency of Figure 4 represents a dispersed map, and the reference saliency of Figure 5 represents a concentrated map. The similarity metrics show consistent performance for the case of Figure 4, meaning they may capture similar properties of DSV. Whereas, the similarity metrics show inconsistent performance for the case of Figure 5, meaning they may focus on different properties of DSV measurement. These examples indicate that the measurement of DSV tends to depend on the dispersion degree of the reference saliency map. Based on these observations, we propose a new similarity

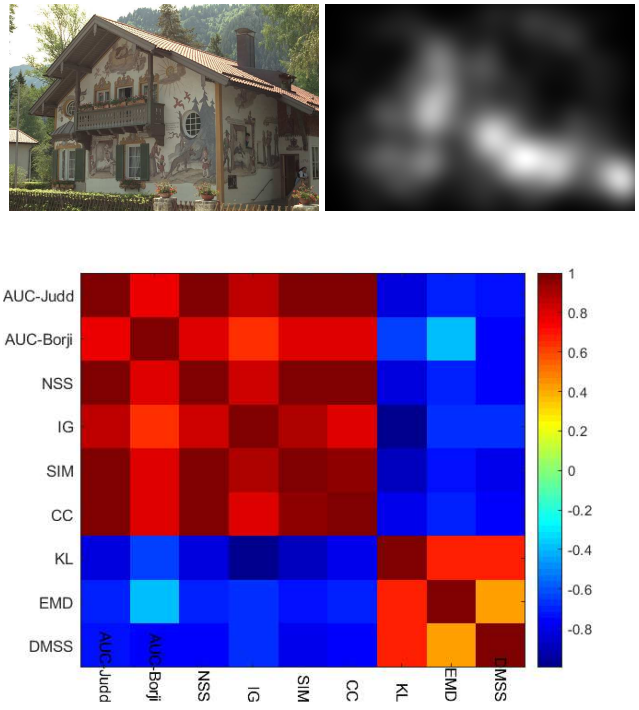



**FIGURE 4.** Correlation matrix visualization of eight saliency similarity metrics and DMSS of all deviated saliency maps from the reference, i.e., "Paint-house".

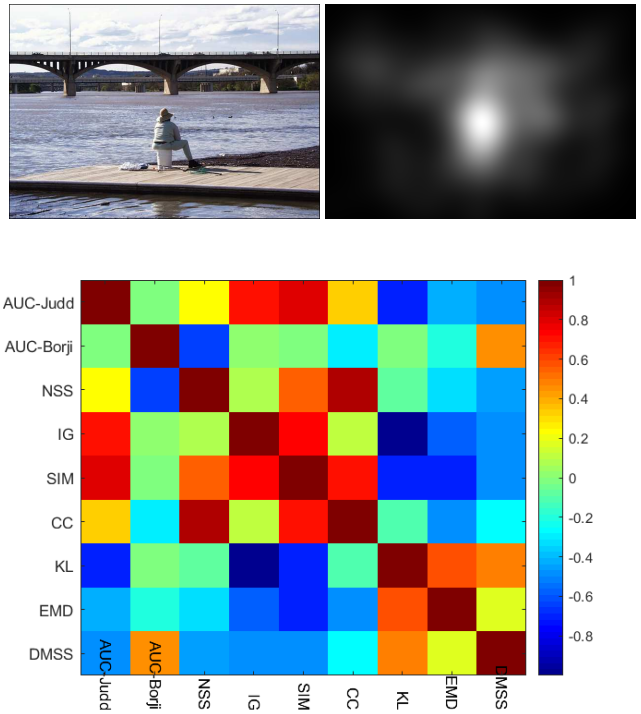



**FIGURE 5.** Correlation matrix visualization of eight saliency similarity metrics and DMSS of all deviated saliency maps from the reference, i.e., "Man-fishing".

metric for assessing the distortion-induced saliency variation (DSV) based on saliency dispersion, which is described in detail below.

## III. THE PROPOSED ALGORITHM

### A. BASIC ALGORITHM FRAMEWORK

First, We propose a new algorithm for quantifying distortion-induced saliency variation (DSV), combining the local and global saliency similarity measures. The principle of the algorithm is to integrate local measure and global measure as demonstrated in the section above; and a modulator based on saliency dispersion is used to control the adaptive measure of saliency similarity. The basic idea of the algorithm is described below.

**Modulator:** Saliency dispersion measure [10] gives a quantitative gauge of the spatial distribution of saliency, i.e., the extent to which a saliency pattern is spread throughout the spatial domain. Based on our observations of the DSV benchmark (as the examples illustrated above), the distortion-induced saliency variation tends to depend on the degree of saliency dispersion: when saliency is concentrated in fewer areas in the spatial domain, the global (i.e., structural characteristics) saliency variation significantly contributes the DSV assessment; when saliency is dispersed throughout the spatial domain, the local (i.e., positional characteristics) saliency variation makes a predominant contribution to the DSV assessment. Hence, the saliency dispersion measure developed in [10] is used here. The method is based on Shannon entropy (note $H$ represents the entropy of a saliency map), using a multilevel approach. More specifically, the saliency map (i.e., $S$) is partitioned at level $P$ into $P \times P$ nonoverlapping blocks of equal size. The multilevel entropy of the saliency maps is defined as:

$$H_{\Sigma}(S) = \frac{1}{P_{\max}} \sum_{P=1}^{P_{\max}} \sum_{B=1}^{N} H(B), \tag{5}$$

where $P_{\max}$ is the finest level of division and $N = P_{\max}^2$, and $B$ runs over each block, as the procedure illustrated in Figure 6. The lower the multilevel entropy, the more compact the saliency map is; otherwise, the higher the multilevel entropy, the more spread-out the saliency map is as illustrated in Figure 6.

**Algorithm formulation:** We now consider how to use the above components for the quantification of the DSV. For the input reference saliency map and the deviated saliency map, local and global saliency similarity metrics are calculated, and the degree of dispersion is calculated for the reference saliency map. The key idea is to reinforce local similarity measure for the spread-out saliency map and global similarity measure for the compact saliency map. We formulate the **m**etric for DSV (MDSV) as a convex optimization problem, and define the metric as:

$$\mathbf{MDSV}(S_d, S_r) = [1 - \sigma(H_{\Sigma}(S_r))]M_1 + \sigma(H_{\Sigma}(S_r))M_2, \tag{6}$$

**TABLE 3.** Performance of state-of-the-art similarity metrics for the measurement of distortion-induced saliency variation (DSV) for "low dispersion" and "high dispersion" saliency maps of the DSV benchmark [17].

| | Criterion | AUC-Judd | AUC-Borji | NSS | IG | SIM | CC | KL | EMD |
|---|---|---|---|---|---|---|---|---|---|
| Low dispersion saliency maps | PLCC | 0.4545 | 0.0518 | 0.4538 | 0.3327 | 0.7394 | 0.7751 | 0.6493 | 0.7136 |
| | SROCC | 0.3755 | 0.0169 | 0.3476 | 0.3064 | 0.6600 | 0.6603 | 0.6770 | 0.7219 |
| | KROCC | 0.2526 | 0.0091 | 0.2395 | 0.2122 | 0.4734 | 0.4953 | 0.5073 | 0.5412 |
| High dispersion sailency maps | PLCC | 0.7518 | 0.3820 | 0.6396 | 0.3157 | 0.5614 | 0.5930 | 0.3379 | 0.5371 |
| | SROCC | 0.7380 | 0.4039 | 0.5686 | 0.2268 | 0.4859 | 0.5495 | 0.3018 | 0.4349 |
| | KROCC | 0.5520 | 0.2281 | 0.4147 | 0.1503 | 0.3458 | 0.4593 | 0.2098 | 0.3023 |

where $M_1$ represents the local metric,$M_2$ represents the global metric, and $\sigma(\cdot)$ represents the revised sigmoid function:

$$\sigma(H_\Sigma(S_r)) = \frac{1}{1+\exp\{-\tau[H_\Sigma(S_r)-T]\}}, \qquad (7)$$

where $T$ is the selected threshold and $\tau$ controls the steepness of the function. In this paper, $T$ and $\tau$ were determined from the saliency maps of the DSV benchmark, using statistical methods and learning-based parameter calibration, which are detailed below.

### B. STATISTICAL PROPERTIES OF SALIENCY SIMILARITY METRICS BASED ON SALIENCY DISPERSION

In order to find suitable components for our proposed convex optimization based DSV metric, we analyse the statistical properties of the state-of-the-art similarity metrics based on saliency dispersion using the DSV benchmark database.

We computed the multilevel entropy of all reference saliency maps contained in the DSV benchmark database [17] using the formula (5). Our previous work showed that the multilevel entropy has the ability to distinguish the dispersion degrees of saliency maps. Figure 6 illustrates the multilevel entropy of two saliency maps of different dispersion degrees: one represents the dispersed saliency and one represents the concentrated saliency.

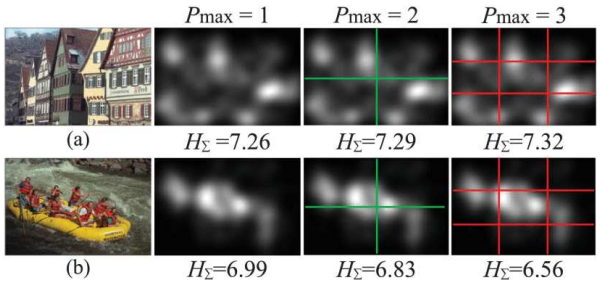



**FIGURE 6.** Calculation of multilevel entropy $H_\Sigma$ for distinguished saliency maps. At each level the saliency map is divided into blocks of equal size. $H_\Sigma$ is found by the sum of the entropy computed at each level of partition. $P_{\max}$ is the level with finest partitioning.

In order to reveal how different similarity metrics respond to different degrees of saliency dispersion, we divided the reference saliency maps in the DSV benchmark into two sets: one set contains saliency maps of high dispersion degrees (i.e., large multilevel entropy); one set contains saliency maps of low dispersion degrees (i.e., small multilevel entropy). The division is made using a threshold (i.e., 7.0126 in our experiment) of the calculated multilevel entropy values for all reference saliency maps. Note rigid thresholding is impossible as dispersion is a relative term, but we have considered the subjective content classification in [26] to assure the division reflects two distinctive degrees of saliency dispersion. This leads to a split of the entire DSV benchmark into "low dispersion" and "high dispersion" saliency maps. To analyse the impact of saliency dispersion on existing similarity metrics, we calculate the performance of these metrics (i.e., based on the PLCC, SROCC, and KROCC between a metric and ground truth DSV) once on the subset of "low dispersion" saliency maps and once on the subset of "high dispersion" saliency maps. Table 3 lists the performance of similarity metrics for "low dispersion" and "high dispersion" saliency maps of the DSV benchmark. As can be seen from Table 3, the metrics' performance differs for different degrees of saliency dispersion. More specifically, for the "low dispersion" saliency maps, globe metrics, i.e., SIM and CC give the best performance amongst all metrics; whereas for the "high dispersion" saliency maps, local metrics, i.e., AUC-Judd and NSS produce the best performance amongst all metrics. This tends to suggest that the way local and global metrics contribute to DSV measurement depends on the degree of saliency dispersion, and therefore their usage should be explicitly considered. Based on above analysis, saliency dispersion measured by multilevel entropy could be used as the modulator to combine the local similarity metric and global similarity metric for a more sophisticated measurement of the distortion-induced saliency variation.

**TABLE 4.** Performance of the Structural Similarity (SSIM) Index, Multiscale Structural Similarity (MS-SSIM) Index, Complex Wavelet Structural Similarity (CW-SSIM) Index, and Information Content Weighted Structural Similarity (IW-SSIM) Index for the measurement of distortion-induced saliency variation (DSV) for saliency maps of the DSV benchmark [17].

| Criterion | SSIM | MS-SSIM | CW-SSIM | IW-SSIM |
|---|---|---|---|---|
| PLCC | 0.7052 | 0.7823 | 0.6718 | 0.7784 |
| SROCC | 0.6351 | 0.7321 | 0.6803 | 0.7791 |
| KROCC | 0.4602 | 0.5689 | 0.4899 | 0.5822 |

### C. PLAUSIBILITY OF USING IMAGE FIDELITY METRICS FOR DSV MEASUREMENT

Since saliency maps are essentially images, we naturally wonder whether the well-established image fidelity metric -

Structural SIMilarity (SSIM) index [27] and its variants [28]–[30] could be used to evaluate the distortion-induced saliency variation. In Table 4, we list the performance of the Structural Similarity (SSIM) index, Multiscale Structural Similarity (MS-SSIM) index, Complex Wavelet Structural Similarity (CW-SSIM) index, and Information Content Weighted Structural Similarity (IW-SSIM) index for the measurement of DSV on the DSV benchmark. As can be seen from the table that SSIM and its variants produce good performance, compared to the overall performance of the state-of-the-art similarity saliency metrics as listed in Table 2. The PLCC values of SSIM and its variants are comparable to the best-performing SIM and CC metrics in Table 2. Since SSIM-based methods measure the structural information of images, they seem to capture the saliency patterns. In addition, the MS-SSIM extract the hierarchical information via a multi-scale approach and the IW-SSIM index obtains sophisticated features based on information theory, therefore, both MS-SSIM and IW-SSIM produce slightly higher performance than SSIM. CW-SSIM index focuses on the linear perturbation around edges and textures, which are less important for saliency patterns. This might cause a relatively lower performance for CW-SSIM.

In summary, these SSIM-based metrics have demonstrated the potential for the DSV measurement, and therefore, can be further investigated for use in our proposed algorithm.

### D. SELECTION OF SUB-MODELS FOR THE PROPOSED DSV METRIC

Our proposed convex optimization framework contains two essential components, i.e., local and global similarity measures. Based on above analysis of existing similarity metrics (including the image fidelity metrics), we now focus on determining which metrics could be included in our proposed algorithm. The details of the selection procedure as well as the reasons of selecting specific metrics are described below.

Let $S_r(i,j)$, $S_d(i,j)$ respectively denote the intensity value of the saliency map of the reference image and its distorted version at the position $(i,j)$, and $F_r(i,j)$, $F_d(i,j)$ are the intensity (binary) value of the fixation map of the reference image and its distorted version at the position $(i,j)$.

**Local saliency similarity:** As already discussed in Section II.B, AUC-Judd, AUC-borji, NSS, and IG are classified as local saliency similarity metrics. As per the overall performance listed in Table 2, AUC-Judd and NSS significantly outperform NSS and IG, therefore, we consider AUC-Judd and NSS as candidates and describe both metrics.

**AUC-Judd** [19]- The ROC (Receiver Operating Characteristic) curve is discretely given by the threshold set $L = \{l_1, \cdots, l_n\}$. For each $l_i, i \in \{1, \cdots, n\}$, its corresponding coordinates are defined as $(\text{TPR}_{l_i}, \text{FPR}_{l_i})$, where $\text{TPR}_{l_i}$, $\text{FPR}_{l_i}$ are the true-positive rate and false-positive rate that can be formulated as follows:

$$\text{TPR}_{l_i} = \frac{tp_{l_i}}{tp_{l_i} + fn_{l_i}}, \text{FPR}_{l_i} = \frac{fp_{l_i}}{fp_{l_i} + tn_{l_i}}, \tag{8}$$

where $tp_{l_i}$, $fp_{l_i}$, $tn_{l_i}$, $fn_{l_i}$ represent the number of the true positives, false positives, true negatives, and false negatives for the corresponding $l_i$-level set classifier, and they can be implemented by level sets of the saliency map of the distorted image $S_d$ and the fixation map of the reference image $F_r$. The AUC-Judd is equal to the area under the ROC curve.

**NSS [21]**- For each fixation location $(i,j)$ of the reference $F_r(i,j)$, the corresponding Normalized Scanpath Saliency (NSS) value for the distorted saliency map (i.e., $S_d(i,j)$) is given by:

$$\text{NSS}(i,j) = \frac{S_d(i,j) - \mu_{S_d}}{\sigma_{S_d}}, \tag{9}$$

where $\mu_{S_d}$, $\sigma_{S_d}$ are the average value and standard deviation of the test saliency map $S_d$. Then the NSS metric for the whole test saliency map $S_d$ is:

$$\text{NSS}(S_d) = \frac{1}{|F_{r,1}|} \sum_{(i,j) \in F_{r,1}} \text{NSS}(i,j), \tag{10}$$

where $F_{r,1} = \{(i,j) | F_r(i,j) = 1\}$.

**Global saliency similarity:** As already discussed in Section II.B, SIM, CC, KL, and EMD are global metrics. Based on the overall performance listed in Table 2, CC gives the best performance amongst all global metrics, therefore, we consider CC as the first candidate. SIM's performance is comparable to (slightly lower than) CC, but as we discussed in Section III.C, the image fidelity metrics are applicable and show good potential for DSV measurement, we hence consider SSIM-based metrics rather than SIM in our algorithm. We also want to consider a third candidate, KL, simply because it has been widely used as part of the loss function in deep learning-based saliency models. We wish to investigate whether KL has its place in our proposed framework. The details of the three candidates are described below.

**CC** [24]- The Pearson's Correlation Coefficient (CC) metric of the test saliency map $S_d$ and the reference saliency map $S_r$ is defined as:

$$\text{CC}(S_d, S_r) = \frac{\sigma(S_d, S_r)}{\sigma(S_d)\sigma(S_r)}, \tag{11}$$

where $\sigma(S_d)$, $\sigma(S_r)$ are the variance of saliency maps $S_d$ and $S_r$, respectively, and $\sigma(S_d, S_r)$ is the covariance between the saliency maps $S_d$ and $S_r$.

**KL** [18] - The Kullback-Leibler divergence (KL) metric of the test saliency map $S_d$ and the reference saliency map $S_r$ is defined as:

$$\text{KL}(S_d, S_r) = \sum_{(i,j) \in S_r} \tilde{S}_r(i,j) \log_2[\epsilon + \frac{\tilde{S}_r(i,j)}{\tilde{S}_d(i,j) + \epsilon}], \tag{12}$$

where $\epsilon$ is a regularization constant.

**SSIM** [27]- The structural similarity (SSIM) index of the test saliency map $S_d$ and the reference saliency map $S_r$ is defined as following mean structural similarity (MSSIM):

$$\text{MSSIM}(S_d, S_r) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \text{SSIM}[S_d(i,j), S_r(i,j)], \tag{13}$$

**TABLE 5.** Performance of different fusion methods LR, MLP, SVM-P, SVM-RFB, M5Rules, RT, RF and our proposed $\text{MDSV}_1$ for combing the sub-models AUC-Judd and CC for DSV measurement on the DSV benchmark [17].

| Criterion | LR | MLP | SVM-P | SVM-RFB | M5Rules | RT | RF | $\text{MDSV}_1$ |
|---|---|---|---|---|---|---|---|---|
| PLCC | 0.3965 | 0.2378 | 0.3894 | 0.3506 | 0.3943 | 0.2063 | 0.3385 | **0.7915** |
| SROCC | 0.3554 | 0.1890 | 0.3528 | 0.2892 | 0.3534 | 0.2052 | 0.3031 | **0.7481** |
| KROCC | 0.2492 | 0.1294 | 0.2455 | 0.1998 | 0.2485 | 0.1387 | 0.2068 | **0.5513** |

where $MN$ is the size of $S_d$ and $S_r$, and $\text{SSIM}[S_d(i,j), S_r(i,j)]$ is formulated as:

$$\frac{(2\mu_{S_d(i,j)}\mu_{S_r(i,j)} + C_1)(2\sigma_{S_d(i,j)S_r(i,j)} + C_2)}{(\mu^2_{S_d(i,j)} + \mu^2_{S_r(i,j)} + C_1)(\sigma^2_{S_d(i,j)} + \sigma^2_{S_r(i,j)} + C_2)}, \quad (14)$$

where local statistics $\mu_{S_d(i,j)}$, $\mu_{S_r(i,j)}$, $\sigma_{S_d(i,j)}$, $\sigma_{S_r(i,j)}$, and $\sigma_{S_d(i,j)S_r(i,j)}$ are estimated by the local window that is defined using an normalized $11 \times 11$ circularly-symmetric Gaussian function $w = \{w_p | p = 1, \cdots, l, \sum_{p=1}^{l} w_p = 1\}$ with the standard deviation of $\frac{3}{2}$. Specifically, these local statistics are estimated as:

$$\mu_{S_d(i,j)} = \sum_{p=1}^{l} w_p S_d(p), \quad (15)$$

$$\mu_{S_r(i,j)} = \sum_{p=1}^{l} w_p S_r(p), \quad (16)$$

$$\sigma_{S_d(i,j)} = [\sum_{p=1}^{l} w_p (S_d(p) - \mu_{S_d(i,j)})^2]^{\frac{1}{2}}, \quad (17)$$

$$\sigma_{S_r(i,j)} = [\sum_{p=1}^{l} w_p (S_r(p) - \mu_{S_r(i,j)})^2]^{\frac{1}{2}}, \quad (18)$$

$$\sigma_{S_d(i,j)S_r(i,j)} = \sum_{p=1}^{l} w_p (S_d(p) - \mu_{S_d(i,j)})(S_r(p) - \mu_{S_r(i,j)}), \quad (19)$$

where $p$ is the $p$-th position of the local window.

In summary, we have now selected two local saliency similarity measures (i.e., AUC-Judd and NSS), three global saliency similarity measures (i.e., CC, KL and SSIM) as the candidate sub-models for our proposed convex optimization framework for quantifying the distortion-induced saliency variation (DSV). We will now investigate the construction of a final DSV metric, based on above basic components.

### *E. SELECTION OF ESSENTIAL PARAMETERS FOR THE PROPOSED DSV METRIC*

In the proposed algorithm framework, as detailed in Section III.A and equation (5)-(7) , there are three essential parameters that are worth discussing. We now give more details below on how these important parameters are determined using empirical methods.

**Maximum level of the multilevel entropy for saliency dispersion measure:** The maximum level of the multilevel entropy as illustrated in Figure 6 is determined by calculating the correlation between the estimated saliency dispersion degree and its ground truth - inter-observer agreement (IOA), as detailed in [10]. By varying the variable, i.e., $P_{\text{max}}$ in equation (5), the correlation (i.e., PLCC) is calculated over the DSV benchmark database as a function of $P_{\text{max}}$. It was found that the PLCC value saturated at the maximum level $P_{\text{max}} = 4$. In order to determine whether there is a significant difference between $P_{\text{max}} = 4$ and the higher maximum levels, a Wilcoxon signed rank test is applied, which is a non-parameter version of $t$-test in the case of non-normality based on the residuals between estimated multilevel entropy $H_\Sigma$ and its ground truth IOA. The test results show that there is no statistically significant difference between $P_{\text{max}} = 4$ and $P_{\text{max}} = 5$, and between $P_{\text{max}} = 4$ and $P_{\text{max}} = 6$ (i.e., in both cases, $p$-values are larger than 0.05). Therefore, we choose the maximum level $P_{\text{max}} = 4$ for the calculation of the multilevel entropy to measure the saliency dispersion of the saliency maps contained in the DSV benchmark.

**Steepness of the sigmoid function:** In our proposed algorithm framework for distortion-induced saliency variation, the parameter $\tau$ is used to control the steepness of the sigmoid function. Inspired by the parameters of the Gaussian probability distribution, we estimate the steepness $\tau$ by the inverse of the unbiased deviation of the dispersion degrees of the reference saliency maps in the DSV benchmark, and the estimated value of $\tau$ is given by the following equation:

$$\hat{\tau} = \frac{1}{\frac{1}{R-1}\sum_{t=1}^{R}[H_\Sigma(S_{r,t}) - \overline{H_\Sigma(S_{r,\cdot})}]^2}, \quad (20)$$

where $R$ is the number of the reference saliency maps in the DSV benchmark, $S_{r,t}$ represents the $t$-th reference saliency map, $\overline{H_\Sigma(S_{r,\cdot})} = \frac{1}{T}\sum_{t=1}^{R} H_\Sigma(S_{r,t})$ is the mean of the dispersion degrees of the reference saliency maps. Based on this statistical method, the steepness of the proposed DSV metric is computed and is equal to 20.62 for the DSV benchmark in this paper.

**Threshold of the sigmoid function:** In order to control the convergence and the robustness of our proposed MDSV framework, we need to determine a suitable threshold $T$ for our MDSV algorithm. Since the modulator of our proposed metric framework is related the dispersion degree of the reference saliency maps, we estimate the threshold as:

$$\hat{T} = \frac{1}{R}\sum_{t=1}^{R} \hat{T}[S_{r,t}], \quad (21)$$

where $\hat{T}[S_{r,t}]$ is the estimated value of the $t$-th reference saliency map in the DSV benchmark, and it can be computed

**TABLE 6.** Performance of different fusion methods LR, MLP, SVM-P, SVM-RFB, M5Rules, RT, RF for combing the sub-models NSS, CC and KL and our proposed $\text{MDSV}_2$ for combing NSS and CC (note KL is not considered as it has little impact on DSV) for DSV measurement on the DSV benchmark [17].

| Criterion | LR | MLP | SVM-P | SVM-RFB | M5Rules | RT | RF | $\text{MDSV}_2$ |
|---|---|---|---|---|---|---|---|---|
| PLCC | 0.7842 | 0.7357 | 0.7796 | 0.7633 | 0.7785 | 0.6592 | 0.7859 | **0.8031** |
| SROCC | 0.7288 | 0.6488 | 0.7164 | 0.6749 | 0.6833 | 0.6057 | 0.7395 | **0.7695** |
| KROCC | 0.5387 | 0.4672 | 0.5243 | 0.4895 | 0.4921 | 0.4321 | 0.5420 | **0.5715** |

by the following formula:

$$\hat{T}[S_{r,t}] = H_{\Sigma}(S_{r,t}) + \frac{\ln\left[\text{dist}(S_{r,t}) - 1\right]}{\hat{\tau}}, \tag{22}$$

where $\hat{\tau}$ is given by equation (20), and $\text{dist}(S_{r,t})$ is computed as:

$$\frac{||\boldsymbol{M}_2(S_{r,t}) - \boldsymbol{M}_1(S_{r,t})||_2^2}{[\boldsymbol{M}_2(S_{r,t}) - \boldsymbol{M}_1(S_{r,t})]^T[\boldsymbol{DMSS}(S_{r,t}) - \boldsymbol{M}_1(S_{r,t})]}, \tag{23}$$

where $\boldsymbol{M}_1(S_{r,t})$ is the metric vector of metric $M_1$ for the test saliency maps originated from the $t$-th reference saliency map in the DSV benchmark, $\boldsymbol{M}_2(S_{r,t})$ is the metric vector of metric $M_2$ for the test saliency maps originated from the $t$-th reference saliency map in the SIQ288 database, $\boldsymbol{DMSS}(S_{r,t})$ is the ground truth vector of metric $DMSS$ for the test saliency maps originated from the $t$-th reference saliency map, and $||\cdot||_2$ is the $L_2$ norm of vector space. For example, if we use the modulator to fuse the local similarity measure AUC-Judd and the global similarity measure CC, the estimated threshold is 4.38.

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

### A. INTERACTIONS OF SUB-MODELS FOR THE CONVEX OPTIMIZATION METHOD

In Section III.D, we have selected potential candidate sub-models (for both local and global saliency similarity measures) to construct our final MDSV metric. Now, we investigate (1) how different combinations of local and global measures and (2) how different fusion methods can affect the final DSV measurement.

Firstly, based on the proposed convex optimization method of equation (6), we define a metric for distortion-induced saliency variation, namely $\text{MDSV}_1$, combing the local saliency similarity measure AUC-Judd and global saliency similarity measure CC. We now compare our combination strategy to other alternatives including linear regression (LR), multi-layer perceptron (MLP) [31], support vector machine with polynomial kernel (SVM-P) [32], SVM with radial basis function (RFB) kernel (SVM-RFB) [32], model tree rules (M5Rules) [33], random tree (RT) [34], and random forest (RF) [35]. Each of above fusion methods was used to combine AUC-Judd and CC to form a DSV measurement (note since they are machine learning-based fusion method, a 10-fold cross-validation was used to generate results for a fair comparison between methods). The performance (i.e., PLCC, SROCC and KROCC) of these DSV measures on the DSV benchmark are listed in Table 5. It can be seen that our $\text{MDSV}_1$ significantly outperforms other DSV measures. Also, it is worth noting that $\text{MDSV}_1$ produces better performance than any of the individual sub-models (see their performance in Table 2), meaning the proposed convex optimization method has proven efficacy.

Secondly, in the literature of computational saliency models, deep learning-based methods (e.g., SAM-VGG and SAM-ResNet [36]) often benefit from a loss function that combines sub-models NSS, CC and KL. By using the above mentioned fusion methods, i.e., LR, MLP, SVM-P, SVM-RFB, M5Rules, RT, and RF, DSV measures can be produced by combining NSS, CC and KL (note KL already shows little impact on DSV measurement (see Table 2), but it is included in these metrics without compromising the performance as these metrics learn the weights of sub-models). Since they are machine learning-based fusion method, a 10-fold cross-validation was used to generate results for a fair comparison between methods. Meanwhile, we can use our proposed convex optimization method to define a new DSV metric, namely $\text{MDSV}_2$, by fusing the local saliency similarity measure NSS and the global saliency similarity measure CC. Note, KL is not considered in our DSV metric as it shows little impact on DSV measurement (see Table 2) and our method is designed to combine only two sub-models. The performance (i.e., PLCC, SROCC and KROCC) of these DSV measures is listed in Table 6, which shows that our proposed $\text{MDSV}_2$ is superior to other DSV measures. This also suggests that our proposed algorithm could be potentially used to improve the loss function of deep learning-based saliency models. Moreover, the performance comparison between $\text{MDSV}_1$ and $\text{MDSV}_2$ has demonstrated the contribution of the modulator (i.e., representing saliency dispersion degree) introduced in our algorithm framework. More specifically, the only difference between $\text{MDSV}_1$ and $\text{MDSV}_2$ is that the former uses AUC-Judd and the latter uses NSS as the local saliency similarity measure. According to Table 7, since AUC-Judd is better than NSS, we would expect that $\text{MDSV}_1$ is better than $\text{MDSV}_2$ in terms of DSV measurement. However, as can be seen in Table 7, $\text{MDSV}_2$ is statistically significantly better than $\text{MDSV}_1$ (i.e., $p < 0.05$ via a $t$-test). This means that the modulator has adaptively combined the sub-models towards optimal performance for DSV measurement.

Thirdly, based on our convex optimization method, we could use SSIM or its variant as the sub-model for global saliency similarity measure. This will produce a new family of DSV metrics, for example, we can define a $\text{MDSV}_3$ that combines the local saliency similarity measure NSS and the

**TABLE 7.** erformance of eight state-of-the-art saliency similarity metric and our proposed metrics, $\text{MDSV}_1$, $\text{MDSV}_2$, and $\text{MDSV}_3$ for DSV measurement on the DSV benchmark [17].

| Criterion | AUC-Judd | AUC-Borji | NSS | IG | SIM | CC | KL | EMD | $\textbf{MDSV}_1$ | $\textbf{MDSV}_2$ | $\textbf{MDSV}_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PLCC | 0.6816 | 0.2787 | 0.6782 | 0.3273 | 0.7475 | 0.7786 | 0.3100 | 0.6121 | 0.7915 | 0.8031 | **0.8060** |
| SROCC | 0.6502 | 0.2106 | 0.6081 | 0.2988 | 0.6962 | 0.7246 | 0.2755 | 0.5502 | 0.7481 | **0.7695** | 0.7692 |
| KROCC | 0.4363 | 0.1373 | 0.4328 | 0.2052 | 0.5063 | 0.5345 | 0.1909 | 0.3832 | 0.5513 | **0.5715** | 0.5709 |
| RMSE | 8.1487 | 10.6952 | 8.1839 | 10.5229 | 7.3978 | 6.9878 | 10.5878 | 8.8064 | 6.8903 | 6.5116 | **6.3541** |

global saliency similarity measure SSIM. The performance of $\text{MDSV}_3$ will be discussed in more detail below.

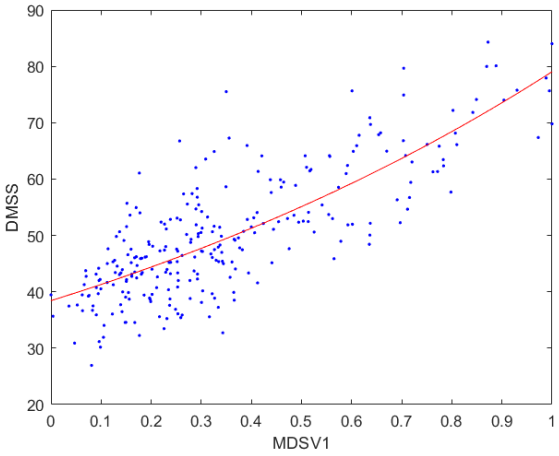



**FIGURE 7.** Scatter plot of the DMSS and our proposed $\text{MDSV}_1$. Curve shows the regression line of nonlinear logistic fitting. X-axis shows the predicted score $\text{MDSV}_1$ and y-axis shows the observers' DMSS.

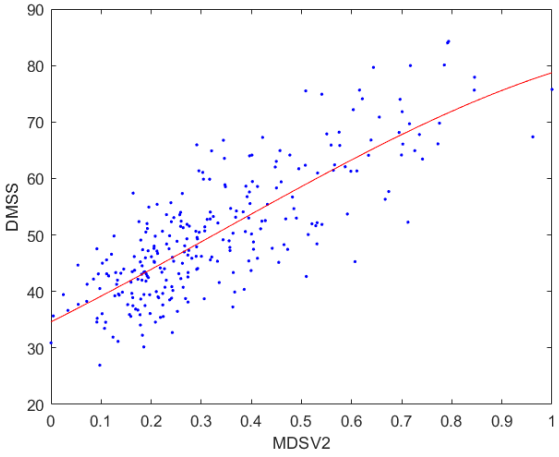



**FIGURE 8.** Scatter plot of the DMSS and our proposed $\text{MDSV}_2$. Curve shows the regression line of nonlinear logistic fitting. X-axis shows the predicted score $\text{MDSV}_2$ and y-axis shows the observers' DMSS.

### B. OVERALL PERFORMANCE

Based on our convex optimization method, we produce three new DSV metrics, $\text{MDSV}_1$, $\text{MDSV}_2$, and $\text{MDSV}_3$ which fuse the local saliency similarity measure and global saliency similarity measure. Now, we give a comprehensive evaluation of these new metrics in comparison to existing saliency similarity metrics. We calculate each new DSV metric for all test (deviated) saliency maps of the DSV benchmark, and quantify the ability of metric outputs to predict the ground truth DMSS scores using PLCC, SROCC, KROCC and RMSE. The overall performance is listed in Table 7 and Figure 7-9 each shows the scatter plot of the DMSS and one of the new DSV metrics. As can be seen, the proposed metrics $\text{MDSV}_1$, $\text{MDSV}_2$, and $\text{MDSV}_3$ outperform existing alternatives for the measurement of DSV. Hypothesis testing (i.e., $t$-test) was conduced and statistical results show that each of our MSDV variant is significant ($p < 0.05$) better than any other alternative metric of the state-of-the-art.

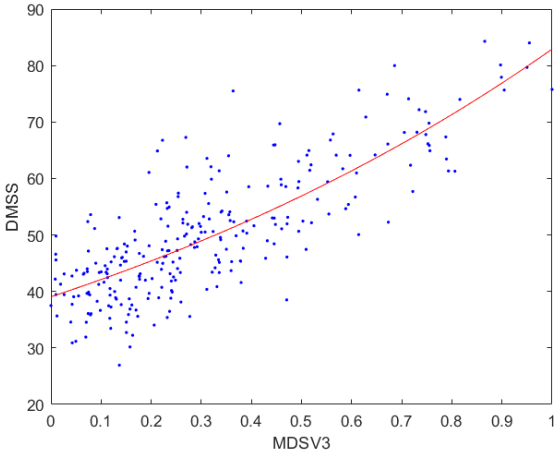



**FIGURE 9.** Scatter plot of the DMSS and our proposed $\text{MDSV}_3$. Curve shows the regression line of nonlinear logistic fitting. X-axis shows the predicted score $\text{MDSV}_3$ and y-axis shows the observers' DMSS.

The superior performance of the proposed metrics demonstrates the importance of taking into account both local and global saliency similarity measures as well as saliency dispersion. It is worth noting that our proposed convex optimization framework is highly adaptable as described in above when $\text{MDSV}_1$, $\text{MDSV}_2$, and $\text{MDSV}_3$ were produced. Different combinations of local and global saliency similarity measures can be easily implemented based on the proposed framework. Based on our current available candidate sub-models, we can continue to construct more variants of MDSV, such as combing AUC-Judd and SSIM. When better sub-models are created in future, they can be easily used to replace current sub-models to produce new MDSV metrics.

**TABLE 8.** Metric output (Mean) for saliency of high (H), medium (M), and low (L) quality images of the DSV benchmark database [17].

| Metrics | Statistics | H | M | L |
|---|---|---|---|---|
| AUC-Judd | Mean | 0.7817 | 0.7757 | 0.7526 |
| | Sig.(p) | H-M:$p > 0.05$ | | M-L:$p < 0.05$ |
| AUC-Borji | Mean | 0.6484 | 0.6454 | 0.6230 |
| | Sig.(p) | H-M:$p > 0.05$ | | M-L:$p < 0.05$ |
| NSS | Mean | 1.2785 | 1.2472 | 1.1006 |
| | Sig.(p) | H-M:$p > 0.05$ | | M-L:$p < 0.05$ |
| IG | Mean | 0.1567 | 0.1348 | -0.0119 |
| | Sig.(p) | H-M:$p > 0.05$ | | M-L:$p < 0.05$ |
| SIM | Mean | 0.8322 | 0.8195 | 0.7661 |
| | **Sig.(p)** | **H-M**:$p < 0.05$ | | **M-L**:$p < 0.05$ |
| CC | Mean | 0.8945 | 0.8776 | 0.7901 |
| | Sig.(p) | H-M:$p > 0.05$ | | M-L:$p < 0.05$ |
| KL | Mean | 0.1830 | 0.1982 | 0.2995 |
| | Sig.(p) | H-M:$p > 0.05$ | | M-L:$p < 0.05$ |
| EMD | Mean | 0.8735 | 0.9864 | 1.2817 |
| | **Sig.(p)** | **H-M**:$p < 0.05$ | | **M-L**:$p < 0.05$ |
| **MDSV$_1$** | Mean | 0.8732 | 0.8515 | 0.7856 |
| | **Sig.(p)** | **H-M**:$p < 0.05$ | | **M-L**:$p < 0.05$ |
| **MDSV$_2$** | Mean | 0.8812 | 0.8546 | 0.7631 |
| | **Sig.(p)** | **H-M**:$p < 0.05$ | | **M-L**:$p < 0.05$ |
| **MDSV$_3$** | Mean | 0.8924 | 0.8463 | 0.7722 |
| | **Sig.(p)** | **H-M**:$p < 0.05$ | | **M-L**:$p < 0.05$ |
| DMSS (Ground truth) | Mean | 45.2084 | 49.0597 | 56.2730 |
| | **Sig.(p)** | **H-M**:$p < 0.05$ | | **M-L**:$p < 0.05$ |

In addition, we analyse an important property of the DSV measurement. One of the significant findings of the previous study [9] is that DSV potentially forms a good basis for image quality prediction. In terms of the association between distorted images and their saliency maps, it is found that the wider the saliency variation relative to the reference (i.e., the larger the difference between the deviated saliency and the original saliency), the higher the distortion in the image (i.e., the lower the image quality). Statistical analysis [17] also revealed that the quality levels are significantly ($p < 0.05$) distinguished by the subjective (ground truth) DMSS. Now, we analyse whether this property also holds for the state-of-the-art similarity metrics and our proposed metrics MDSV$_1$, MDSV$_2$, and MDSV$_3$. Table 8 lists the statistical analysis for the similarity metrics. It can be seen that all metrics can distinguish medium and low quality levels, but most of them (except for SIM, EMD and our proposed MDSV$_1$, MDSV$_2$, and MDSV$_3$) fail in distinguishing the high and medium quality levels. Overall, results demonstrate that our proposed metrics give the best performance for the measurement of the distortion-induced saliency variation.

## V. CONCLUSION

Following up on our previous work of building a benchmark for distortion-induced saliency variation (DSV), in this paper, we proposed a novel convex optimization framework for quantifying DSV. We have investigated eight state-of-the-art saliency similarity metrics and their predictive performance for DSV. From the statistical analyses, we found that the performance of these metrics for DSV highly depends on the dispersion degree of the saliency maps. Based on the findings, we therefore proposed an algorithm framework based on the convex optimization method to combine local and global saliency similarity measures and using saliency dispersion measure as the modulator. We have produced three new DSV metrics, MDSV$_1$, MDSV$_2$, and MDSV$_3$ based off our proposed framework. Experimental results show that these new DSV metrics significantly outperform existing metrics in quantifying distortion-induced saliency variation (DSV). Our proposed framework is highly adaptable, and can be easily updated given better-performing local and global sub-models are created. Going forward, the proposed DSV metric will be used to facilitate the development of advanced saliency models that can predict saliency of distorted images in various image quality related applications.

## REFERENCES


[1] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric," in 2007 IEEE International Conference on Image Processing, vol. II. IEEE, 2007, pp. 169–172.

[2] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," IEEE transactions on Circuits and Systems for Video Technology, vol. 21, no. 7, pp. 971–982, 2011.

[3] F. Röhrbein, P. Goddard, M. Schneider, G. James, and K. Guo, "How does image noise affect actual and predicted human gaze allocation in assessing image quality?" Vision research, vol. 112, pp. 11–25, 2015.

[4] Y. Guo, Z. Li, Y. Liu, G. Yan, and M. Yu, "Video object extraction based on spatiotemporal consistency saliency detection," IEEE Access, vol. 6, pp. 35 171–35 181, 2018.

[5] A. Bruno, F. Gugliuzza, R. Pirrone, and E. Ardizzone, "A multi-scale colour and keypoint density-based approach for visual saliency detection," IEEE Access, vol. 8, pp. 121 330–121 343, 2020.

[6] J. Li, X. Feng, and H. Fan, "Saliency consistency-based image re-colorization for color blindness," IEEE Access, vol. 8, pp. 88 558–88 574, 2020.

[7] U. Engelke, H. Kaprykowsky, H. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," IEEE Signal Processing Magazine, vol. 28, no. 6, pp. 50–59, 2011.

[8] L. Zhang, Y. Shen, and H. Li, "Vsi: A visual saliency-induced index for perceptual image quality assessment," IEEE Transactions on Image processing, vol. 23, no. 10, pp. 4270–4281, 2014.

[9] W. Zhang, A. Borji, Z. Wang, P. Le Callet, and H. Liu, "The application of visual saliency models in objective image quality assessment: A statistical evaluation," IEEE transactions on neural networks and learning systems, vol. 27, no. 6, pp. 1266–1278, 2015.

[10] W. Zhang, R. R. Martin, and H. Liu, "A saliency dispersion measure for improving saliency-based image quality metrics," IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 6, pp. 1462–1466, 2017.

[11] L. Tang, Q. Wu, W. Li, and Y. Liu, "Deep saliency quality assessment network with joint metric," IEEE Access, vol. 6, pp. 913–924, 2018.

[12] T.-J. Liu, "Study of visual quality assessment on pattern images: Subjective evaluation and visual saliency effects," IEEE Access, vol. 6, pp. 61 432–61 444, 2018.

[13] X. Xu, B. Shi, Z. Gu, R. Deng, X. Chen, A. S. Krylov, and Y. Ding, "3d no-reference image quality assessment via transfer learning and saliency-guided feature consolidation," IEEE Access, vol. 7, pp. 85 286–85 297, 2019.

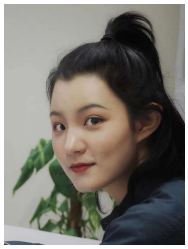
[14] C. Shi and Y. Lin, "Full reference image quality assessment based on visual salience with color appearance and gradient similarity," IEEE Access, vol. 8, pp. 97 310–97 320, 2020.

[15] W. Zhang and H. Liu, "Toward a reliable collection of eye-tracking data for image quality research: challenges, solutions, and applications," IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2424–2437, 2017.

[16] ——, "Siq288: A saliency dataset for image quality research," in 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 2016, pp. 1–6.

[17] L. Leveque, W. Zhang, and H. Liu, "Subjective assessment of image quality induced saliency variation," in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 1024–1028.
[18] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?" IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 3, pp. 740–757, 2018.
[19] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, "Mit saliency benchmark," 2015.
[20] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study," IEEE Transactions on Image Processing, vol. 22, no. 1, pp. 55–69, 2012.
[21] R. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," Vision research, vol. 45, no. 18, pp. 2397–2416, 2005.
[22] M. Kümmerer, T. Wallis, and M. Bethge, "Information-theoretic model comparison unifies saliency metrics," Proceedings of the National Academy of Sciences, vol. 112, no. 52, pp. 16 054–16 059, 2015.
[23] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," International journal of computer vision, vol. 40, no. 2, pp. 99–121, 2000.
[24] O. L. Meur, P. L. Callet, and D. Barba, "Predicting visual fixations on video based on low-level visual features," Vision research, vol. 47, no. 19, pp. 2483–2498, 2007.
[25] O. Pele and M. Werman, "A linear time histogram metric for improved sift matching," in European conference on computer vision. Springer, 2008, pp. 495–508.
[26] 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 Processing Letters, vol. 20, no. 4, pp. 355–358, 2013.
[27] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.
[28] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003, vol. 2, 2003, pp. 1398–1402 Vol.2.
[29] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2385–2401, 2009.
[30] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Transactions on Image Processing, vol. 20, no. 5, pp. 1185–1198, 2011.
[31] R. Fabrice and C.-G. Brieuc, "Functional multi-layer perceptron: a non-linear tool for functional data analysis," Neural Networks, vol. 18, no. 1, pp. 45–60, 2005.
[32] C.-C. Chang and C.-J. Lin, "Libsvm - a library for support vector machines," 2001, the Weka classifier works with version 2.82 of LIBSVM. [Online]. Available: http://www.csie.ntu.edu.tw/c̃jlin/libsvm/
[33] H. Geoffrey, H. Mark, and F. Eibe, "Generating rule sets from model trees," in Twelfth Australian Joint Conference on Artificial Intelligence. Springer, 1999, pp. 1–12.
[34] A. David, "The continuum random tree. i," The Annals of Probability, vol. 19, no. 1, pp. 1–28, 1991. [Online]. Available: http://www.jstor.org/stable/2244250
[35] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[36] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "Predicting human eye fixations via an lstm-based saliency attentive model," IEEE Transactions on Image Processing, vol. 27, no. 10, pp. 5142–5154, 2018.

PENGFEI GUO received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2015. He is currently an Associate Professor with the School of Computing Science, Zhongkai University of Agriculture and Engineering, Guangzhou, China. He has been the visiting scholar of Cardiff University. His research interests include machine learning, image processing, computer vision and image quality assessment.



XIN ZHAO is currently a Ph.D. student at Cardiff University, Cardiff, U.K. She received her BSc in Computer Science from Cardiff University in 2019. She was a visiting scholar at Konstanz University, Konstanz, Germany in 2020. Her research interests include eye tracking, image quality assessment and machine learning.

DONGQING WU received the M.S. degree in computer graphics from the Institute of Industrial Design and Graphics, South China University of Technology, Guangzhou, China, in 2005, and the Ph.D. degree in mechanical engineering from the School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou, China, in 2019. He is currently a visiting scholar in the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China, and an associate professor with the School of Computational Science, Zhongkai University of Agriculture and Engineering, Guangzhou, China. His current research interests include neural networks, geometry algorithm, and pattern recognition.

LEI ZHANG received the B.S. degree in Computer Science and Technology from Henan Polytechnic University, Jiaozuo China in 2006, the M.S. degree in Computer Application from South China Normal University, Guangzhou, China in 2009. He is currently a experimentalist with the College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou, China. His recent research interests include the areas of intelligent information system of agriculture, data mining, software engineering and applications to agricultural engineering.

DELU ZENG received his Ph.D. degree in electronic and information engineering from South China University of Technology, China, in 2009. He is now a full professor in the School of Mathematics in South China University of Technology, China. He has been the visiting scholar of Columbia University, University of Oulu, University of Waterloo. He has been focusing his research in applied mathematics and its interdisciplinary applications. His research interests include numerical calculations, applications of partial differential equations, optimizations, machine learning and their applications in image processing, and data analysis.

HANTAO LIU received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands in 2011. He is currently an Associate Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. He is an Associate Editor of the IEEE Transactions on Human-Machine Systems and the IEEE Transactions on Multimedia.