IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 May 2023, accepted 29 May 2023, date of publication 6 June 2023, date of current version 14 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3283344

RESEARCH ARTICLE

CUDAS: Distortion-Aware Saliency Benchmark

XIN ZHAO[1], JIANXUN LOU[1], XINBO WU[1], YINGYING WU[1], LUCIE LÉVÉQUE[2], XIAOCHANG LIU[3], PENGFEI GUO[4], (Member, IEEE), YIPENG QIN[1], HANHE LIN[5], DIETMAR SAUPE[6], AND HANTAO LIU[1]

[1]School of Computer Science and Informatics, Cardiff University, CF24 4AG Cardiff, U.K.
[2]LS2N-UMR 6004 CNRS, Nantes University, 44000 Nantes, France
[3]School of Materials, Sun Yat-sen University, Guangzhou 510275, China
[4]School of Computational Science, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
[5]School of Science and Engineering, University of Dundee, DD1 4HN Dundee, U.K.
[6]Department of Computer and Information Science, University of Konstanz, 78457 Konstanz, Germany

Corresponding authors: Jianxun Lou (louj2@cardiff.ac.uk) and Xinbo Wu (wux37@cardiff.ac.uk)

This work was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through TRR 161 (Project A05) under Project 251654672.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the School of Computer Science and Informatics Research Ethics Committee, Cardiff University.

ABSTRACT Visual saliency prediction remains an academic challenge due to the diversity and complexity of natural scenes as well as the scarcity of eye movement data on where people look in images. In many practical applications, digital images are inevitably subject to distortions, such as those caused by acquisition, editing, compression or transmission. A great deal of attention has been paid to predicting the saliency of distortion-free pristine images, but little attention has been given to understanding the impact of visual distortions on saliency prediction. In this paper, we first present the CUDAS database - a new distortion-aware saliency benchmark, where eye-tracking data was collected for 60 pristine images and their corresponding 540 distorted formats. We then conduct a statistical evaluation to reveal the behaviour of state-of-the-art saliency prediction models on distorted images and provide insights on building an effective model for distortion-aware saliency prediction. The new database is made publicly available to the research community.

INDEX TERMS Eye-tracking, saliency, distortion, image quality, deep learning.

I. INTRODUCTION

Visual attention is a primary mechanism of the human visual system (HVS) that enables selecting the most relevant information in a visual field [1], [2], [3], [4]. It allows humans to focus their perceptual-cognitive capability on certain important stimuli in the scene while ignoring irrelevant stimuli. Visual attention consists of both bottom-up (content-driven, stimulus-driven) mechanism and top-down (task-driven, experience-driven) mechanism [5], [6]. The former refers to the ability of the HVS to unconsciously detect the salient stimuli in the visual scene, which potentially benefits many research fields including multimedia, computer vision, and healthcare [7], [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Long Xu.

The past few decades have witnessed significant progress in visual saliency modelling. The advances in this field are largely attributed to the creation of benchmark databases for human eye movements, which provide insights on functional mechanism of attention. Popular databases include MIT300 [9], CAT2000 [10], and SALICON [11]. The MIT300 dataset is widely used to benchmark the performance of visual saliency models and contains 300 indoor and outdoor images. The CAT2000 dataset offers a diverse range of image content, featuring 4,000 images across 20 scene categories, such as action, art, and cartoon, with 200 images per category. SALICON dataset uses mouse clicks rather than eye trackers to collect human visual attention. SALICON is one of the largest saliency datasets available in the literature, comprising 10,000 training images, 5,000 validation images, and 5,000 test images.

Using these benchmark databases, many computational models have been proposed to automatically predict visual saliency [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34]. These models take various approaches to model saliency and generate a saliency map that indicates conspicuousness of different scene locations. The so-called traditional method relies on extracting low-level visual features, such as colour, intensity and orientation and combining these features to form a saliency map. This method requires modelling the functionality of the HVS, which remains a challenging task. The other method is based on deep neural networks, where models are trained to predict a saliency map. Due to the advances of deep learning techniques, this method can achieve good performance.

In many practical applications, images are inevitably subject to distortions such as those caused by acquisition, editing, compression and transmission. Previous research has shown that the saliency map of a pristine image differs from that of a distorted format of the image; and that the degree of the difference depends on the type and level of distortion [35], [36]. Modelling saliency in the context of image distortion is highly beneficial for various vision computing applications, e.g., image quality assessment (IQA) [37] where an accurate saliency map of a distorted image is required to optimise the objective IQA metrics [38]. Unfortunately, there is still a lack of benchmark databases in the literature to comprehensively address the problem of saliency modelling for distorted images. Pioneering work was conducted in [39], where the SIQ288 database was created using eye-tracking to understand the impact of distortions on human gaze. This work focused on developing an experimental methodology for reliably collecting eye movement data for images of varying degrees and types of distortion. However, the database remains limited in terms of diversity in image content, for example.

In order to drive this line of research forward, we create a new distortion-aware saliency benchmark, namely **Cardiff University Distortion-Aware Saliency (CUDAS) database**. Based on the new benchmark, we analyse the behaviour of state-of-the-art saliency models in predicting saliency of distorted images. The contributions of this work are detailed below.

- We apply a reliable experimental methodology in [39] to conduct a large-scale eye-tracking study on distorted images, resulting in the largest-of-its-kind distortion-aware saliency database, namely CUDAS. It consists of 60 pristine (undistorted) high-quality and high-resolution source images from 10 different categories of visual content. These source images were degraded using three types of distortion and each at three levels of distortion. This gives a set of 600 stimuli including originals. A total of 96 subjects were recruited to participate in the eye-tracking study in a fully controlled lab environment, which ensures the reliability of the ground truth saliency maps of the CUDAS database.
- We conduct an exhaustive analysis on the behaviour of 20 state-of-the-art saliency models, including both traditional and deep learning-based models, on the CUDAS database. We investigate the effect of distortion level and distortion type on the performance of these saliency model. This provides valuable insights for model selection in practical application scenarios.
- We investigate plausible solutions towards building an effective deep learning-based model for predicting saliency of distorted images. We provide quantitative evidence on the impact of transfer learning as well as different network architectures on the model's predictive power. This provides a foundation for further research on computational saliency modelling.

II. CUDAS: EYE-TRACKING STUDY

A. STIMULI

We use the same set of stimuli of the CUID database [40], which contains both pristine images and their corresponding distorted formats. There are 60 high-quality and high-resolution (1920 × 1080 pixels) pristine images. One important feature of the CUID database is that the stimuli are content-rich and categorised into 10 different natural scene categories, including ACT (Action), BNW (Black and White), CGI (Computer-Generated Imagery), IND (Indoor), OBJ (Object), ODM (Outdoor Manmade), ODN (Outdoor Natural), PAT (Pattern), POT (Portrait), and SOC (Social). The content and scene categories are illustrated in Fig. 1. The distorted images are generated by simulating three different types of distortion including contrast change (CC), JPEG compression (JPEG), and motion blur (MB). By varying the magnitude of distortion, three distinctive levels (i.e., Q1, Q2 and Q3) of perceived quality are created, representing low-level perceptible but not noticeable distortion, medium-level annoying distortion, and high-level very annoying distortion. As a result, a set of 600 stimuli was yielded.

B. EYE-TRACKING EXPERIMENT

As per the findings in [39], to ensure the reliability of eye-tracking data collected for the co-occurrence of pristine images and their distorted formats, a between-subjects method [41] combined with appropriate control mechanisms must be applied in the experiment. This is done to avoid subject biases [6] due to stimulus repetition - same scene content (with varying degrees of quality) repeatedly shown to the subject. By use of the experimental protocol proposed in [39], we divided the stimuli into six partitions of 100 images each. Each partition contained a mixture of all types and all levels of distortion, and at most two repeated formats of the same scene content. Following the design concept of a between-subjects experiment, we had to ask six different groups of subjects to each view one of these partitions of stimuli. To this end, 96 subjects being 48 females and 48 males with ages ranging from 19 to 55 years old were recruited (informed consent was obtained) in our eye-tracking study. We divided the subjects

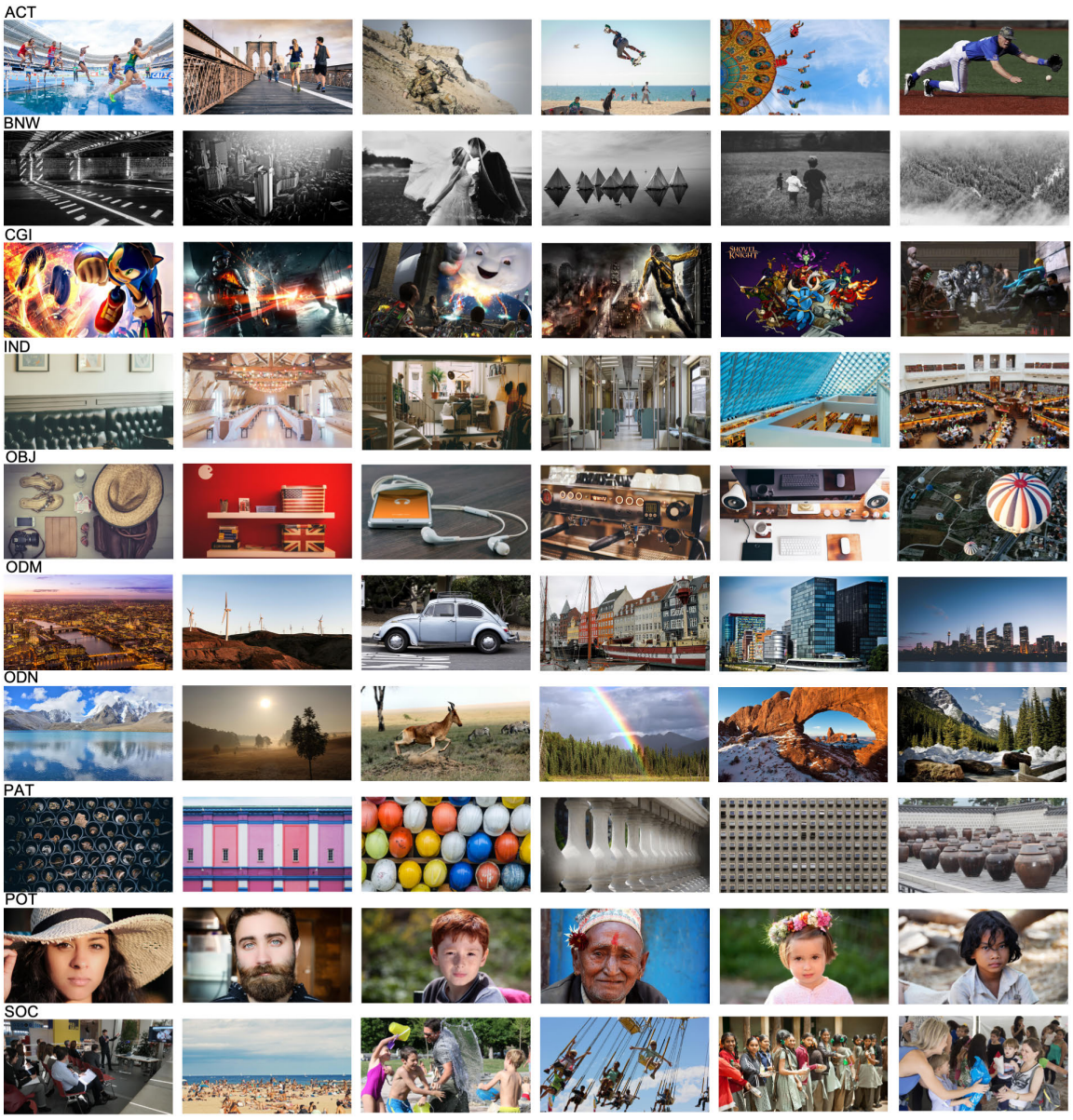
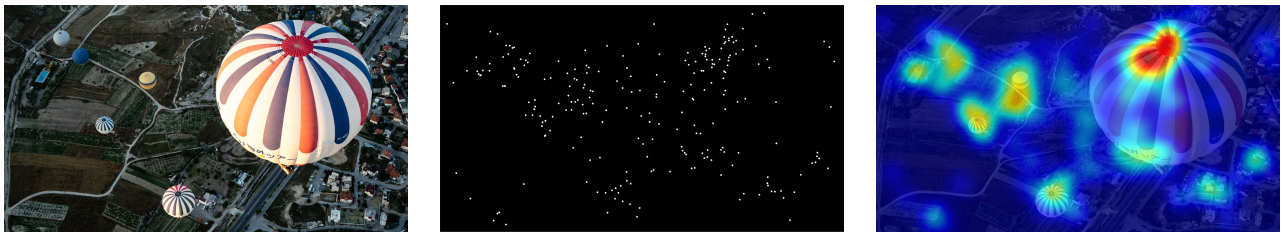


FIGURE 1. Illustration of the 60 source images (6 pristine images × 10 scene categories) contained in the CUID database [40]. From top to bottom, the categories are ACT (Action), BNW (Black and White), CGI (Computer-Generated Imagery), IND (Indoor), OBJ (Object), ODM (Outdoor Manmade), ODN (Outdoor Natural), PAT (Pattern), POT (Portrait), and SOC (Social).



(a) Pristine reference (b) Fixation map (c) Saliency map

FIGURE 2. Illustration of an example of saliency data contained in the new CUDAS database.

into six groups of 16 subjects each (including 8 females and 8 males); and assigned one group to view one partition of stimuli. This means each subject had to view 100 images from the entire dataset. By doing this gives a sample size of

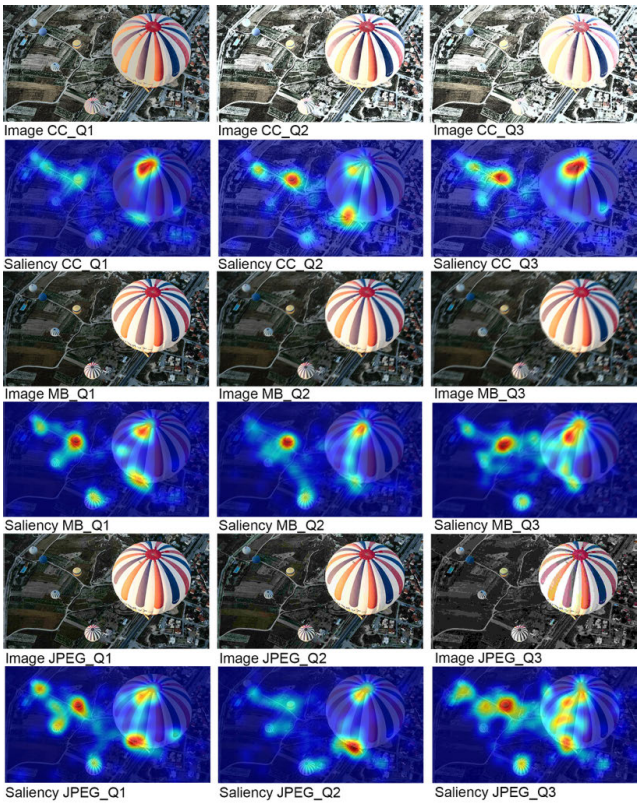


FIGURE 3. Illustration of an example of saliency maps created for 9 different distorted formats of a pristine image (see Fig.2) contained in the new CUDAS database. The database contains three different types of distortion including contrast change (CC), JPEG compression (JPEG), and motion blur (MB); and three distinctive levels (i.e., Q1, Q2 and Q3) of perceived quality.

16 subjects per test image, which meets the requirement [39] for generating a reliable saliency map for a natural image. For each subject's trial, we also broke the session into two sub-sessions and added a "washout" period of 5 hours to reduce the carry-over effects [39]. Each sub-session contained no stimulus repetition and stimuli were presented to each subject in a random order.

Following the International Telecommunication Union (ITU) standards [42], we set up a standard office environment at the Visual Computing Laboratory at Cardiff University for the conduct of our eye-tracking experiment. The fully controlled viewing environment e.g., constant ambient light ensured consistent experimental conditions throughout the entire experiment. A 19-inch LCD monitor (native resolution 1920×1080 pixels) was used to display stimuli to subjects. The viewing distance was maintained approximately 60 cm. Eye movements were captured using a non-invasive SensoMotoric Instrument (SMI) Red-m advanced eye tracking system with a sampling rate of 250 Hz. Subjects were asked to view the stimuli in a natural manner using the instruction, "view the image as you normally would". In our study, each image was displayed for five seconds, followed by a mid-gray screen of two seconds.

C. SALIENCY MAP GENERATION

We follow the standard approach [39] to generate saliency maps from the eye-tracking data. In this study, a ground truth fixation was rigorously defined by the SMI BeGaze Analysis Software using an established dispersal and duration based algorithm [43]. For each image, fixations extracted from the raw eye-tracking data collected over all 16 subjects are converted to a saliency map:

$$\mathrm{SM}(x, y) = \sum_{i=1}^{N} \exp\left[-\frac{(x_i - x)^2 + (y_i - y)^2}{\sigma^2}\right], \quad (1)$$

where SM (x, y) represents the saliency map; $(x_i,\ y_i)$ represents the coordinates of i-th fixation; N is the number of total fixations; and σ is the standard deviation of the Gaussian ($\sigma = 45$ pixels determined as per [44] in our study). In creating a saliency map (also known as fixation density map), a Gaussian patch is added to each fixation location to simulate the foveal vision (i.e., 2° of visual angle) of the human visual system (HVS). The intensity of the resulting saliency map is linearly normalised to the range [0, 1]. Now, the Cardiff University Distortion-Aware Saliency (CUDAS) database is generated. Fig. 2 shows the fixation map and saliency map (visualised as a heat map) for one of the pristine images in the CUDAS database. Fig.3 illustrates the saliency maps generated for 9 different distorted formats of the pristine image.

III. CUDAS: PERFORMANCE BENCHMARKING AND EVALUATION OF STATE-OF-THE-ART SALIENCY MODELS

A. EVALUATION FRAMEWORK

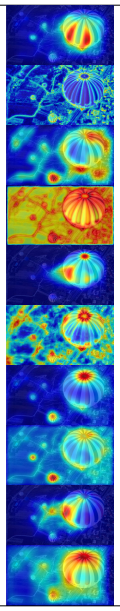
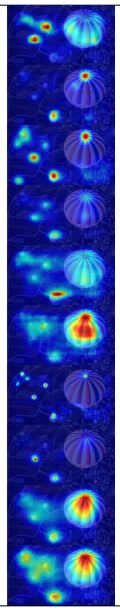
1) SALIENCY MODELS

In this study, we aim to benchmark the performance of state-of-the-art saliency models and reveal their behaviour on distorted images contained in the CUDAS database. We selected 10 traditional models [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] and 10 deep learning-based models [22], [23], [24], [25], [26], [27], [28], [29], [30]. These models have been proven rather effective in predicting saliency of pristine images [45]; however, their performance on distorted images remains largely unexplored. To this end, we evaluate (1) the overall performance of state-of-the-art saliency models on the CUDAS database; (2) the impact of distortion (both level and type) on the performance of these saliency models. The brief descriptions of the selected saliency models are included in Table 1. Note to make a fair comparative study in this section, all models were implemented without re-calibration (for traditional models) or re-training (for deep learning-based models) with the CUDAS database and they are assumed to be readily applicable for saliency prediction.

2) SALIENCY EVALUATION METRICS

To measure the performance of computational saliency models against the ground truth saliency generated from eye-tracking, some saliency evaluation metrics have been proposed in the literature [46]. It should be noted that these

TABLE 1. Descriptions of the selected saliency models and their predicted saliency maps. The saliency maps are generated by each model for a pristine image (see "Pristine reference" in Fig.2) and a distorted image (see "Image CC_Q3" in Fig.3).

Traditional Saliency Models			Deep Learning-Based Saliency Models		
Model Name & Description	Pristine	Distorted	Model Name & Description	Pristine	Distorted
• **GBVS** [12] constructs a Markov chain for feature maps that are extracted using a similar approach to *ITTI*.			• **iSEEL** [22] is based on inter-image similarities and an ensemble of Extreme Learning Machines (ELM); and includes an image feature transformer using VGG-16 architecture.		
• **Torralba** [13] is based on a Bayesian framework to compute image saliency and global-context features in parallel.			• **ML-Net** [23] combines features extracted at different levels of a VGG16 network and can learn a custom priority (i.e., fixation bias).		
• **ITTI** [14] models the salient locations by using proto-objects form volatile units of visual information that can be accessed by selective attention.			• **Deepgaze II** [24] adopts a fixed VGG-19 network to provide features and a point-wise non-linearity read-out network for saliency prediction.		
• **AIM** [15] is based on the principle of maximising self-information from a scene.			• **SalGAN** [25] is a Generative Adversarial Network (GAN) with a VGG-based encoder-decoder-style generator and a discriminator.		
• **FES** [16] detects saliency by using a centre-surround approach based on estimating saliency of local feature contrast in a Bayesian framework.			• **CASNetII**: [26] includes a component for image context awareness and two parallel encoder networks with different input spatial sizes.		
• **CIWaM** [17] unifies chromatic assimilation and chromatic contrast effects based on spatial frequency and contrast surrounding energy assumptions.			• **DVA** [27] is based on a skip-layer network structure to capture saliency information with a different granularity from different network levels.		
• **AWS** [18] is based on contextual adaptation mechanisms to ensure the invariance of visual system behaviour in response to optical variability.			• **EML-NET** [28] is a scalable system to leverage multiple powerful deep CNN models to better extract visual features for saliency prediction.		
• **CovSal** [19] uses region covariances of simple features extracted from image scene patches to estimate a saliency map.			• **SAM-VGG** [29] combines CNN (VGG-16) with a Convolutional Long Short-Term Memory network (ConvLSTM) and can learn custom priorities.		
• **LDS** [20] estimates image saliency by learning a set of discriminative subspaces to separate salient targets and distractors.			• **SAM-ResNet** [29] consists of an identical architecture to SAM-VGG expect for replacing VGG-16 in SAM-VGG with ResNet-50.		
• **UHM** [21] is a multi-scale hierarchical saliency model, which utilizes both local and global saliency pipelines.			• **MSI-Net** [30] contains multiple convolution layers at different dilution rates to capture multi-scale features.		

metrics capture different properties of saliency evaluation, and "*specific tasks and applications may call for a different choice of metrics*" [46]. In our study, we focus on saliency in the context of image distortion, which is highly relevant for applications such as image compression, transmission, enhancement and re-targeting. It is suggested in [46] that in those applications where it is important to evaluate the relative importance of different image regions, metrics like Correlation Coefficient (CC) and Similarity (SIM) are the best fit. The studies conducted by Yang et al. [47] and Li et al. [48] suggest that CC and SIM are the only two metrics which are in close agreement with human subjective assessments of saliency maps; and that other metrics are not "perception-based" and therefore cannot measure the perceptual relevance of the computational saliency maps. Therefore, CC and SIM metrics are used in our study to benchmark and evaluate saliency models on the CUDAS database. Hereby we give brief descriptions of CC and SIM metrics below, letting PM and SM be the predicted saliency map and ground truth saliency map, respectively.

Correlation Coefficient (CC): CC measures the linear correlation between PM and SM:

$$\mathrm{CC}(\mathrm{PM}, \mathrm{SM}) = \frac{\mathrm{cov}(\mathrm{PM}, \mathrm{SM})}{\sigma_{\mathrm{PM}} \times \sigma_{\mathrm{SM}}}, \tag{2}$$

where σ_{PM}, σ_{SM} denote the variance of PM and SM, and cov(PM, SM) denotes the covariance of the two saliency maps. The range of CC value is between -1 and 1. When CC is close to 1 or -1, the two saliency maps PM and SM are highly similar. When CC is close to 0, the two saliency maps PM and SM largely differ.

Similarity (SIM): SIM is also called histogram intersection, which measures the similarity between the two saliency maps PM and SM when they are rendered as a normalised histogram of pixel intensities (i.e., denoted as PM_i or SM_i). SIM is calculated as:

$$\mathrm{SIM}(\mathrm{PM}, \mathrm{SM}) = \sum_i \min(\mathrm{PM}_i, \mathrm{SM}_i), \tag{3}$$

and

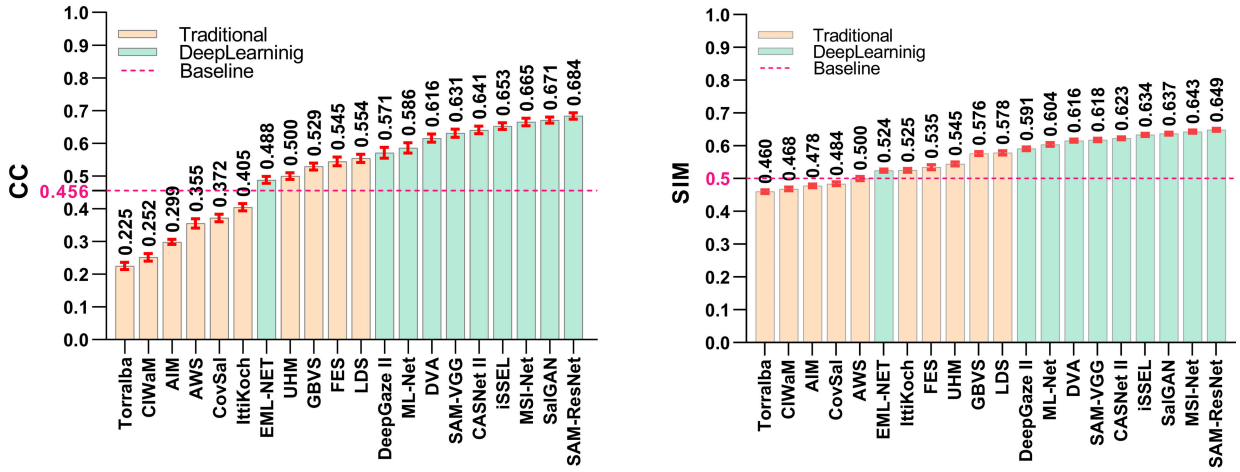
$$\sum_i \mathrm{PM}_i = \sum_i \mathrm{SM}_i = 1, \tag{4}$$

where i represents the index of the histogram. The range of SIM value is between 0 to 1. A higher SIM value indicates a higher degree of agreement between the two saliency maps PM and SM.

B. PERFORMANCE EVALUATION

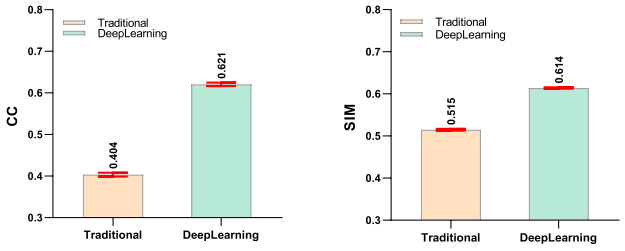
1) OVERALL PERFORMANCE ON CUDAS

For each saliency model, its performance is quantified by calculating CC (or SIM) between the predicted saliency map PM and the ground truth saliency map SM over all stimuli contained in the CUDAS database. Fig.4 shows the ranking results of the models' performance in terms of CC and SIM, respectively. The *baseline* performance, as defined in [39], is used to gauge the effectiveness of a saliency model. The baseline assumes that the centre of an image is the most salient region, which can be simply modelled by stretching a symmetric Gaussian to fit the aspect ratio of the image. It can be seen from Fig.4 that all deep learning-based models are above the baseline performance and half of traditional



(a) (b)

FIGURE 4. Illustration of the performance of 20 state-of-the-art saliency models, including traditional and deep learning-based models on the CUDAS database. The performance is measured by two saliency evaluation metrics, (a) CC and (b) SIM. The baseline assumes that the centre of an image is the most salient region [45]. The error bars indicate a 95 % confidence interval.



(a) (b)

FIGURE 5. Average performance of traditional versus deep learning-based saliency models on the CUDAS database. The performance is measured by two saliency evaluation metrics (a) CC and (b) SIM. The error bars indicate a 95% confidence interval.

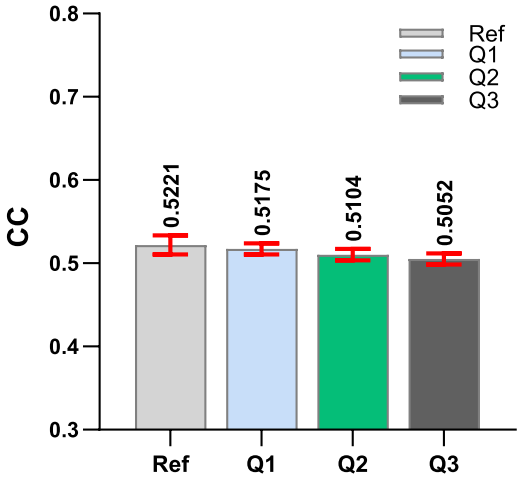


FIGURE 6. Average performance (measured by the CC metric) of all models on different distortion levels including the pristine reference on the CUDAS database. Ref, Q1, Q2 and Q3 represent reference, high quality, medium quality, and low quality images, respectively. The error bars indicate a 95% confidence interval.

models are below the baseline performance; and that SAM-ResNet, MSI-Net and SalGAN are consistently ranked top 3 in both CC and SIM rankings. To statistically verify whether the performance of deep learning-based saliency models is significantly higher than that of traditional models, we perform hypothesis testing by selecting CC (or SIM) as the dependent variable and the categorical model type (traditional versus deep learning) as the independent variable. The Mann-Whitney U test [49] is performed (due to evidence of non-normality as per the Shapiro-Wilk test), and the results ($p < 0.05$) show that CC (or SIM) of the deep learning-based models is statistically significantly higher than that of the traditional models. Fig.5 illustrates the mean CC and SIM for traditional and deep learning-based saliency models.

2) IMPACT OF DISTORTION STRENGTH

In the CUDAS database, pristine images are degraded with three distinctive distortion levels Q1, Q2 and Q3. We want to evaluate whether and to what extent the strength of distortion can affect the performance of saliency models. Fig.6 shows the average performance in terms of CC (note SIM exhibits the same trend as CC, and therefore, is not shown here) of all models on different distortion levels including the pristine reference. It tends to indicate that the stronger the distortion (i.e., the lower the image quality), the lower performance of the saliency models. To statistically verify the impact of distortion strength, we perform hypothesis testing by selecting the CC metric as the dependent variable and distortion level as the independent variable. The Mann-Whitney U test is performed, and the results show that the saliency models perform significantly better on higher quality images than on lower quality images. A statistical significance ($p < 0.05$)

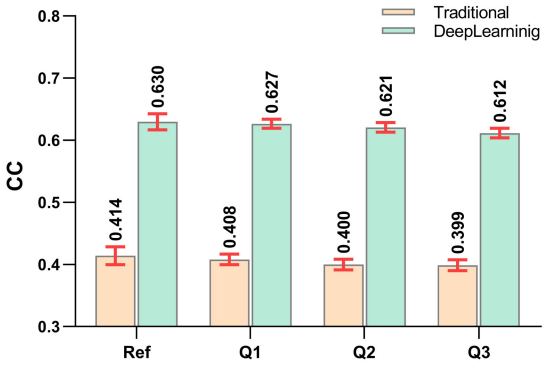


FIGURE 7. Average performance (measured by the CC metric) of traditional versus deep learning-based saliency models on different distortion levels including the pristine reference on the CUDAS database. Ref, Q1, Q2 and Q3 represent reference, high quality, medium quality, and low quality images, respectively. The error bars indicate a 95% confidence interval.

is found between the following variables: Ref>Q3, Q1>Q3, and Q2>Q3. The evidence suggests that predicting saliency of distorted images is more challenging than pristine images for current saliency models, especially there is a significant drop in performance when the distortion becomes stronger (i.e., at Q3 level).

Also, the impact of distortion strength holds the same trend when separating the saliency models into traditional and deep learning-based types. Fig.7 illustrates the average performance (measured by the CC metric) for all deep learning-based models and all traditional models separately at four levels of image quality (i.e., reference, Q1, Q2 and Q3). The results of hypothesis testing (i.e., Mann-Whitney U test) show that saliency models perform significantly better ($p < 0.05$) on higher quality images than on lower quality images, regardless of whether it is a deep learning-based model or a traditional model.

3) IMPACT OF DISTORTION TYPE

In addition, we evaluate whether and to what extent distortion type including contrast change (CC), JPEG compression (JPEG), and motion blur (MB) has an impact on the performance of saliency models. Fig.8 shows the average performance in terms of CC (note SIM exhibits the same trend as CC, and therefore, is not shown here) of all models on different distortion types. Hypothesis testing was performed by selecting CC as the dependent variable and distortion type as the independent variable. The Mann-Whitney U test results show that the performance of saliency models for contrast change (CC) is significantly lower ($p < 0.05$) than the pristine reference. However, the impact of JPEG compression (JPEG) and motion blur (MB) on the performance of saliency models is negligible ($p > 0.05$). By separating deep learning-based models from traditional models, as shown in Fig.9, we have found that there is no significant difference ($p > 0.05$ by Mann-Whitney U test) in performance when using deep learning-based models to predict saliency of images with different distortion types. This suggests that

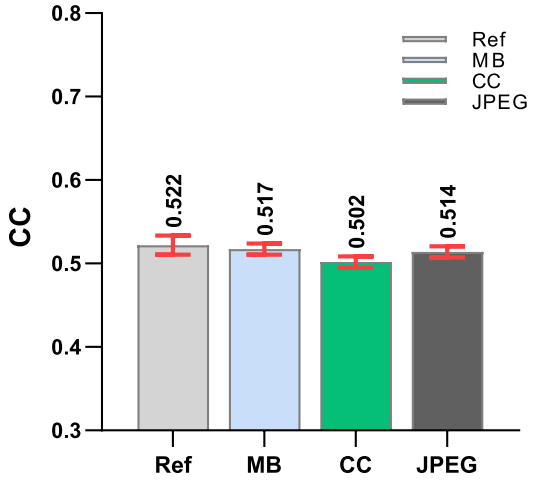


FIGURE 8. Average performance (measured by the CC metric) of all models on different distortion types on the CUDAS database. Ref, MB, CC and JPEG represent reference, motion blur (MB), contrast change (CC), and JPEG compression (JPEG), respectively. The error bars indicate a 95% confidence interval.

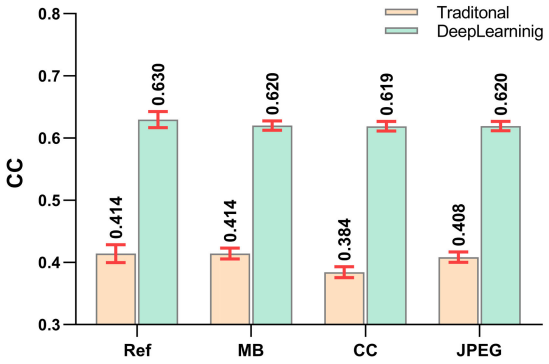


FIGURE 9. Average performance (measured by the CC metric) of traditional versus deep learning-based saliency models on different distortion types on the CUDAS database. Ref, MB, CC and JPEG represent reference, motion blur (MB), contrast change (CC), and JPEG compression (JPEG), respectively. The error bars indicate a 95% confidence interval.

traditional models are more sensitive to different types of distortion, such as contrast change in images; however, the performance of deep learning-based saliency models is consistent over all distortion types.

In summary, deep learning-based models give significantly better performance than the traditional models independent of distortion strength and distortion type. For deep learning-based models, the current challenge lies in handling varying degrees of distortion in a consistent manner.

IV. INSIGHTS ON DEEP LEARNING MODELLING

The evidence above shows the potential of deep learning-based models in solving the problem of saliency prediction in the context of image distortion. Now, we provide further practicalities towards a deep learning-based solution for saliency modelling.

A. IMPACT OF TRANSFER LEARNING

Unlike traditional models, deep learning-based models are data-driven and can readily benefit from fine-tuning on

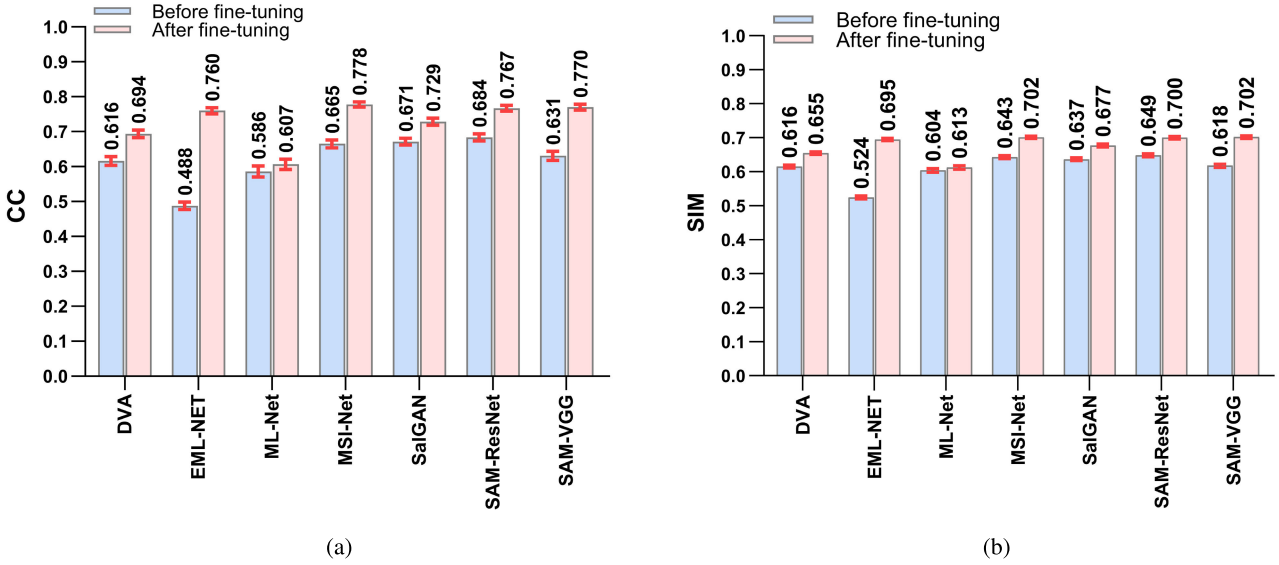


FIGURE 10. Illustration of the performance (measured by the (a) CC and (b) SIM metrics) of deep learning-based saliency models before and after fine-tuning on the CUDAS database. The error bars indicate a 95 % confidence interval.

the target data. To solve our specific problem of predicting saliency of distorted images, one way is to investigate whether fine-tuning the existing models on the new CUDAS database will give good results. To this end, we conduct experiments using the deep learning-based models (in Section III) that are fine-tunable, including ML-Net, SalGAN, DVA, EML-NET, SAM-VGG, SAM-ResNet, and MSI-Net. More specifically, these models were first loaded with the pre-trained parameters on SALICON dataset (note this is a conventional operation for all deep learning-based saliency models), then fine-tuned on the CUDAS dataset. The hyper-parameters in fine-tuning for each model were set based on the recommendations in the literature and our empirical evidence of achieving the lowest loss value in the validation set during fine-tuning. To obtain comprehensive results, k-fold cross-validation ($k = 6$) is applied for each model's fine-tuning trial, where the CUDAS dataset is divided into six non-overlapping sets of equal size. To prevent data leakage, we ensure that there is no shared scene content between sets: (1) first, the 60 source images (see Fig. 1) are divided into six sets of 10 images each (i.e., each set contains one of the columns of the 10 $\times$ 6 grid gallery as illustrated in Fig.1); (2) second, within each set, its 10 source images and their corresponding 90 distorted images (10 scenes $\times$ 3 distortion types $\times$ 3 distortion levels) form the samples. At each run of the k-fold cross-validation, one set is kept as a test set, one as a validation set, and the remaining four sets altogether are used as the training set. By doing so, the model is evaluated on the test set of unseen samples. It should be noted that each run of the k-fold cross-validation is independently conducted, without parameter sharing between runs. We report the average performance of k=6 times test results. Fig. 10 illustrates the performance (measured by CC and SIM) of saliency models before and after fine-tuning on the CUDAS database. It can be seen from Fig. 10 that fine-tuning deep learning-based models on the CUDAS dataset can significantly boost (i.e, $p < 0.05$ by Wilcoxon signed-rank test) their performance. This implies that when predicting saliency of distorted images, fine-tuning models on the ground truth saliency data of distorted images is essential and that training models on the pristine image saliency data only will not provide optimal solution. This means collecting more eye-tracking databases in the context of image distortion will facilitate research in this area.

Now, we have produced a new leaderboard on the CUDAS database, showing the ability of saliency models in handling distorted images. On the other hand, the leaderboard of MIT benchmark [50] shows the ability of saliency models in handling pristine images. It is valuable to cross compare the two leaderboards to guide the selection of models for specific applications. To this end, we calculate the Kendall rank correlation coefficient (KRCC) between two different performance leaderboard rankings. Note since the variant of EML-NET implemented by the MIT Benchmark is different from the variant (provided in the original publication) applied in our study, EML-NET is excluded from this comparison. The KRCC is 0.07 for the CC metric and 0.47 for the SIM metric, indicating a weak correlation and the performance of saliency models on these two application domains is inconsistent. Table 2 shows the comparison of performance rankings for the CUDAS and MIT benchmarks. It can be seen that models that perform well on the MIT benchmark are not necessarily good models for predicting saliency of distorted images in the CUDAS benchmark, and vice versa. This implies that the selection of saliency models in the context of image distortion cannot rely on the existing

TABLE 2. Comparison of performance rankings of deep learning-based models on the CUDAS (distortion-aware image saliency) and MIT (pristine image saliency) benchmarks.

Model Names	CC (CUDAS / MIT)	SIM (CUDAS / MIT)
SAM-VGG	1st / 6th	1st / 3rd
SAM-ResNet	2nd / 2nd	2nd / 2nd
MSI-Net	3rd / 1st	4th / 1st
SalGAN	4th / 3rd	3rd / 4th
DVA	5th / 5th	5th / 5th
ML-Net	6th / 4th	6th / 6th

TABLE 3. Comparison of performance of deep learning-based models with versus without a machine attention mechanism (i.e., attentive versus non-attentive) on the CUDAS database. The *p*-value represents the results of the Mann-Whitney U test for statistical significance.

Metric	Attentive models	Non-Attentive models	p-value
CC ↑	0.7688	0.7136	< 0.05
SIM ↑	0.7011	0.6683	< 0.05

saliency benchmarks which were developed for pristine images.

B. IMPACT OF NETWORK ARCHITECTURES

1) THE USE OF MACHINE ATTENTION MECHANISM

As shown in Table 2, the Top2 fine-tuned deep learning-based models on both CC and SIM metrics are SAM-VGG and SAM-ResNet. Unlike other models, they both utilise the machine attention mechanism in their decoders. More specifically, SAM-VGG and SAM-ResNet combine a fully convolutional network with a recurrent convolutional network (i.e., long short-term memory network), endowed with a spatial attention mechanism. We compare the models with versus without a machine attention mechanism (i.e., attentive versus non-attentive), as the results shown in Table 3. The hypothesis testing (i.e., Mann-Whitney U test) results show that the average performance of attentive models is significantly higher ($p < 0.05$) than that of non-attentive models. This may be attributed to the fact that attentive models use network modules with long-range modelling capabilities to better simulate the processes of the human visual system, and that this mechanism could be especially beneficial for tasks closely related to human perception, e.g., viewing natural scenes in the occurrence of image distortions.

2) THE USE OF DIFFERENT BACKBONES

The choice of backbone formation for deep learning-based models determines the effectiveness of the image feature extraction for saliency prediction. There are two different types of backbone formations used in the selected deep learning-based saliency models, being single-stream encoder and two-stream encoder. EML-NET uses a two-stream encoder consisting of two deep backbones; while other models adopt a single-stream CNN-based backbone.

TABLE 4. Comparison of performance of deep learning- based models using single-stream encoder versus two-stream encoder on the CUDAS database. The *p*-value represents the results of the Mann-Whitney U test for statistical significance.

Metric	Single-stream encoder	Two-stream encoder	p-value
CC ↑	0.7242	0.7603	<0.05
SIM ↑	0.6747	0.6950	<0.05

We compare the use of these two different backbone formations in predicting saliency of distorted images. Table 4 shows the performance of single-stream encoder versus two-stream encoder on the CUDAS database. Hypothesis testing (i.e., Mann-Whitney U test) results show that saliency models based on a two-stream encoder achieve significantly better ($p < 0.05$) performance than models based on a single-stream encoder. This suggests that using backbones with strong representation capabilities is highly beneficial for saliency prediction in the context of image distortion.

V. DISCUSSION

It should be noted that in this paper we focus on the analysis of saliency in the context of image distortion and the perceptual relevance of computational saliency models. For our specific application, CC and SIM are the most appropriate saliency evaluation metrics, as elaborated in Section III-A.2). However, the CUDAS database can also be used as a standalone saliency benchmark, evaluating saliency models for various applications using different saliency evaluation metrics. For example, CUDAS can be used to benchmark models for salient object detection in a noisy environment, where AUC, KL-Div, and IG (centerbias as the baseline map [45]) are appropriate metrics for detection applications as they penalise target detection failures [46]. To facilitate benchmarking, we hereby provide the results of evaluating state-of-the-art saliency models (i.e. models used in Table 2) based on widely used saliency metrics including CC (Correlation Coefficient), SIM (Similarity), AUC-Judd (Area under ROC Curve-Judd) [51], AUC-Borji (Area under ROC Curve-Borji) [52], KL-Div (Kullback-Leibler divergence) [45], NSS (Normalized Scanpath Saliency) [53], and IG (Information Gain) [54]. Table 5 lists the model performance results on the CUDAS database, where divergent rankings across metrics are evident. This further supports the earlier finding in the literature, "specific tasks and applications may also call for a different choice of metrics" [46].

Also, it should be noted that in this paper we use the same set of stimuli of the previously published image quality assessment database CUID [40], where the distortion simulation is limited to three different types. In the CUID database, these distortion types were chosen to reflect three distinctive image impairments commonly occurring in real-world applications: CC affects the colours of images, JPEG yields local artifacts, and MB causes global distortions. As illustrated in Fig. 2 and 3, different distortion types tend to impact the

TABLE 5. Performance of state-of-the-art saliency models measured by CC, SIM, AUC-Borji, AUC-Judd, KL-Div, NSS, and IG metrics on the CUDAS database. The best result for each metric is highlighted in bold.

Model Names	CC↑	SIM↑	AUC-Borji↑	AUC-Judd↑	KL-Div↓	NSS↑	IG↑
SAM-VGG	0.631	0.618	0.702	0.788	1.087	1.354	-0.550
SAM-ResNet	**0.684**	**0.649**	**0.778**	**0.798**	0.984	**1.392**	-0.461
MSI-Net	0.665	0.643	0.776	0.793	0.626	1.347	0.030
SalGAN	0.671	0.637	0.766	0.796	0.824	1.379	-0.191
DVA	0.616	0.616	0.768	0.782	**0.538**	1.243	**0.147**
ML-Net	0.586	0.604	0.724	0.769	0.676	1.284	0.019

saliency distribution in different ways. However, as already thoroughly discussed in [39], the actual impact is a function of image content, distortion type and distortion level. In this paper, we focus on the impact of distortions on saliency prediction models, as the findings revealed in Section III-B. These advances call for larger benchmarks and more distortion types.

VI. CONCLUSION

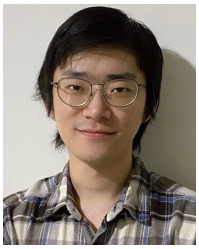
In this paper, we have presented a new distortion-aware saliency benchmark - CUDAS database - to facilitate saliency modelling in the context of image distortion. The CUDAS database contains 60 high-quality, high-resolution, and content-rich pristine images and their corresponding 540 distorted images of varying degrees of perceived quality. We have conducted an exhaustive evaluation to benchmark the performance of 20 state-of-the-art saliency models on the CUDAS database. We have found that deep learning-based models give significantly better performance than traditional models; but there is still room for improvement in terms of handing different degrees of distortion in images. Based on the new benchmark, we shed light on deep learning-based saliency modelling in the context of image distortion, including the impact of transfer learning, use of machine attention mechanism and choice of network backbone formation. Future work will focus on developing a deep-learning based model which can reliably predict saliency of images of varying degrees of perceived quality.

REFERENCES

[1] W. James, *The Principles of Psychology*. Cambridge, MA, USA: Harvard Univ. Press, 1890.
[2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Dec. 1998.
[3] J. Wu, Y. Liu, L. Li, and G. Shi, "Attended visual content degradation based reduced reference image quality assessment," *IEEE Access*, vol. 6, pp. 12493–12504, 2018.
[4] Y. Fang, W. Lin, B. Lee, C. Lau, Z. Chen, and C. Lin, "Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 187–198, Feb. 2012.
[5] C. E. Connor, H. E. Egeth, and S. Yantis, "Visual attention: Bottom-up versus top-down," *Current Biol.*, vol. 14, no. 19, pp. R850–R852, Oct. 2004.
[6] A. G. Greenwald, "Within-subjects designs: To use or not to use?" *Psychol. Bull.*, vol. 83, no. 2, pp. 314–320, Mar. 1976.
[7] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 55–69, Jan. 2013.
[8] L. Lévêque, H. Bosmans, L. Cockmartin, and H. Liu, "State of the art: Eye-tracking studies in medical imaging," *IEEE Access*, vol. 6, pp. 37023–37034, 2018.
[9] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," MIT Comput. Sci. Artif. Intell. Lab (CSAIL), Cambridge, MA, USA, Tech. Rep. MIT-CSAIL-TR-2012-001, Jan. 2012.
[10] A. Borji and L. Itti, "CAT2000: A large scale fixation dataset for boosting saliency research," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshop Future Datasets*, Feb. 2015, pp. 1–4.
[11] M. Jiang, S. Huang, J. Duan, and Q. Zhao, "SALICON: Saliency in context," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1072–1080.
[12] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. 19th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Cambridge, MA, USA: MIT Press, 2006, pp. 545–552.
[13] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search.," *Psychol. Rev.*, vol. 113, no. 4, pp. 766–786, Oct. 2006.
[14] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.*, vol. 19, no. 9, pp. 1395–1407, Nov. 2006.
[15] N. Bruce and J. Tsotsos, "Attention based on information maximization," *J. Vis.*, vol. 7, no. 9, p. 950, Mar. 2010.
[16] H. Rezazadegan Tavakoli, E. Rahtu, and J. Heikkilä, "Fast and efficient saliency detection using sparse sampling and kernel density estimation," in *Image Analysis* (Lecture Notes in Computer Science), vol. 6688. Berlin, Germany: Springer, 2011, pp. 666–675.
[17] X. Otazu, C. A. Parraga, and M. Vanrell, "Toward a unified chromatic induction model," *J. Vis.*, vol. 10, no. 12, pp. 1–24, Oct. 2010.
[18] A. Garcia-Diaz, V. Leboran, X. R. Fdez-Vidal, and X. M. Pardo, "On the relationship between optical variability, visual saliency, and eye fixations: A computational approach," *J. Vis.*, vol. 12, no. 6, p. 17, Jun. 2012.
[19] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *J. Vis.*, vol. 13, no. 4, p. 11, Mar. 2013.
[20] S. Fang, J. Li, Y. Tian, T. Huang, and X. Chen, "Learning discriminative subspaces on random contrasts for image saliency analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1095–1108, May 2017.
[21] R. Tavakoli and J. Laaksonen, "Bottom-up fixation prediction using unsupervised hierarchical models," in *Computer Vision—ACCV*, vol. 10116. New York, NY, USA: Springer-Verlag, 2017, pp. 364–379.
[22] H. R. Tavakoli, A. Borji, J. Laaksonen, and E. Rahtu, "Exploiting inter-image similarity and ensemble of extreme learners for fixation prediction using deep features," *Neurocomputing*, vol. 244, pp. 10–18, Jun. 2017. [Online]. Available: https://arxiv.org/abs/1610.06449
[23] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "A deep multi-level network for saliency prediction," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 3488–3493.
[24] M. Kümmerer, T. S. A. Wallis, L. A. Gatys, and M. Bethge, "Understanding low- and high-level contributions to fixation prediction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4799–4808.

[25] J. Pan, C. C. Ferrer, K. McGuinness, N. E. O'Connor, J. Torres, E. Sayrol, and X. G. I. Nieto, "SalGAN: Visual saliency prediction with generative adversarial networks," 2018, *arXiv:1701.01081*.
[26] S. Fan, Z. Shen, M. Jiang, B. L. Koenig, J. Xu, M. S. Kankanhalli, and Q. Zhao, "Emotional attention: A study of image sentiment and visual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7521–7531.
[27] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2368–2378, May 2018.
[28] S. Jia and N. D. B. Bruce, "EML-NET: An expandable multi-layer NETwork for saliency prediction," *Image Vis. Comput.*, vol. 95, Mar. 2020, Art. no. 103887.
[29] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "Predicting human eye fixations via an LSTM-based saliency attentive model," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5142–5154, Oct. 2018.
[30] A. Kroner, M. Senden, K. Driessens, and R. Goebel, "Contextual encoder–decoder network for visual saliency prediction," *Neural Netw.*, vol. 129, pp. 261–270, Sep. 2020.
[31] J. Lou, H. Lin, D. Marshall, D. Saupe, and H. Liu, "TranSalNet: Towards perceptually relevant visual saliency prediction," *Neurocomputing*, vol. 494, pp. 455–467, Jul. 2022.
[32] M. Tliba, M. A. Kerkouri, B. Ghariba, A. Chetouani, A. Çöltekin, M. S. Shehata, and A. Bruno, "SATSal: A multi-level self-attention based architecture for visual saliency prediction," *IEEE Access*, vol. 10, pp. 20701–20713, 2022.
[33] F. Qi, C. Lin, G. Shi, and H. Li, "A convolutional encoder–decoder network with skip connections for saliency prediction," *IEEE Access*, vol. 7, pp. 60428–60438, 2019.
[34] 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. 121330–121343, 2020.
[35] X. Zhao, H. Lin, P. Guo, D. Saupe, and H. Liu, "Deep learning vs. traditional algorithms for saliency prediction of distorted images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 156–160.
[36] L. Leveque, W. Zhang, and H. Liu, "Subjective assessment of image quality induced saliency variation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1024–1028.
[37] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.
[38] 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 Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1266–1278, Jun. 2016.
[39] W. Zhang and H. Liu, "Toward a reliable collection of eye-tracking data for image quality research: Challenges, solutions, and applications," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2424–2437, May 2017.
[40] L. Lévêque, J. Yang, X. Yang, P. Guo, K. Dasalla, L. Li, Y. Wu, and H. Liu, "CUID: A new study of perceived image quality and its subjective assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 116–120.
[41] G. Keren and C. Lewis, *A Handbook for Data Analysis in the Behavioral Sciences: Volume 1: Methodological Issues Volume 2: Statistical Issues*, 1st ed. London, U.K.: Psychology Press, 1993.
[42] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT.500-11, 1974.
[43] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proc. Symp. Eye Tracking Res. Appl. (ETRA)*. New York, NY, USA: Association for Computing Machinery, 2000, pp. 71–78.
[44] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 970–982, Jan. 2000.
[45] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba. *Mit Saliency Benchmark*. Accessed: 2023. [Online]. Available: http://saliency.mit.edu/
[46] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 740–757, Mar. 2019.
[47] X. Yang, F. Li, and H. Liu, "A measurement for distortion induced saliency variation in natural images," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021.
[48] J. Li, C. Xia, Y. Song, S. Fang, and X. Chen, "A data-driven metric for comprehensive evaluation of saliency models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 190–198.
[49] A. Field, *Discovering Statistics Using IBM SPSS Statistics*. Newbury Park, CA, USA: Sage, 2013.
[50] M. Kümmerer, Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba. *Mit/tübingen Saliency Benchmark*. Accessed: 2023. [Online]. Available: https://saliency.tuebingen.ai/
[51] N. Riche, M. Duvinage, M. Mancas, B. Gosselin, and T. Dutoit, "Saliency and human fixations: State-of-the-art and study of comparison metrics," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1153–1160.
[52] A. Borji, H. R. Tavakoli, D. N. Sihite, and L. Itti, "Analysis of scores, datasets, and models in visual saliency prediction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 921–928.
[53] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vis. Res.*, vol. 45, no. 18, pp. 2397–2416, Aug. 2005.
[54] M. Kümmerer, T. S. A. Wallis, and M. Bethge, "Information-theoretic model comparison unifies saliency metrics," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 52, pp. 16054–16059, Dec. 2015.

XIN ZHAO received the Bachelor of Science degree from the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K., in 2019. She is currently pursuing the Ph.D. degree with Cardiff University. She has been a Visiting Scholar with Konstanz University, Konstanz, Germany. Her research interests include eye-tracking, image quality assessment, and human fixations on distortion images.



JIANXUN LOU received the B.Eng. degree from Central South University, Changsha, China, in 2018, and the M.S. degree from Cardiff University, Cardiff, U.K., in 2020, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Informatics.

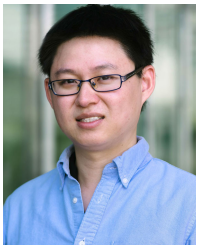
XINBO WU received the B.Eng. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2018, and the M.S. degree from Cardiff University, Cardiff, U.K., in 2020, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Informatics.

YINGYING WU received the M.Sc. degree in data science and analytics from Cardiff University, U.K., in 2020, where she is currently pursuing the Ph.D. degree with the School of Computer Science and Informatics. Her research interests include image data analysis, human visual perception, and machine learning.

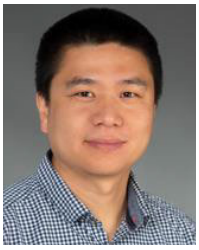
LUCIE LÉVÉQUE received the M.Eng. degree in cognitive engineering from the National Superior School of Cognitics, Bordeaux, France, in 2015, the M.Sc. degree in biomedical imaging from the University of Angers, France, in 2015, and the Ph.D. degree from the School of Computer Science and Engineering, Cardiff University, U.K., in 2019. She is currently a Postdoctoral Researcher with the Nantes Laboratory of Digital Sciences (LS2N), Nantes University, France. Her research interests include human–computer interaction, computer vision, visual perception and attention, and medical imaging. She is also the Vice Chair of the Video Quality Experts Group (VQEG) on Quality Assessment for Health Applications, and part of the Organizing Committee of the ACM International Conference on Interactive Media Experiences (IMX) 2023.

XIAOCHANG LIU is currently pursuing the bachelor's degree with the School of Materials, Sun Yat-sen University, China. Her research interests include mathematical modeling and data analytics.

PENGFEI GUO (Member, IEEE) received the Ph.D. degree from the South China University of Technology, China, in 2015. He was a Visiting Scholar with Cardiff University, U.K. He is currently an Associate Professor with the School of Computing Science, Zhongkai University of Agriculture and Engineering, China. His research interests include computer vision and image quality assessment.

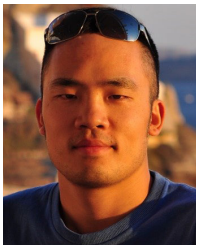


YIPENG QIN received the B.Sc. degree in electrical engineering from Shanghai Jiao Tong University, China, and the Ph.D. degree from the National Centre for Computer Animation (NCCA), Bournemouth University, U.K. He was a Postdoctoral Research Fellow with the Visual Computing Center (VCC), King Abdullah University of Science and Technology (KAUST), Saudi Arabia. He is currently a Lecturer with the School of Computer Science and Informatics, Cardiff University, U.K. His current research interests include deep learning, computer vision, computer graphics, and human–computer interaction (HCI), with a focus on generative modeling and visual content creation.



HANHE LIN received the Ph.D. degree from the Department of Information Science, University of Otago, New Zealand, in 2016. From 2016 to 2021, he was a Postdoctoral Researcher with the Department of Computer and Information Science, University of Konstanz, Germany, where he was working on project A05 (visual quality assessment) of SFB-TRR 161, funded by the German Research Foundation (DFG). He is currently a Lecturer in computing with the University of Dundee, U.K. His research interests include image processing, computer vision, machine learning, deep learning, and visual quality assessment. He serves as a member for the technical program committee or a reviewer for a number of conferences/journals, such as QoMEX, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and IEEE TRANSACTIONS ON IMAGE PROCESSING.

DIETMAR SAUPE was born in Bremen, Germany, in 1954. He received the Dr.rer.nat. degree in mathematics from the University of Bremen, Germany, in 1982. From 1985 to 1993, he was an Assistant Professor with the Department of Mathematics, first with the University of California at Santa Cruz, Santa Cruz, CA, USA, and then with the University of Bremen, resulting in his habilitation. From 1993 to 1998, he was a Professor of computer science with the University of Freiburg, Germany. He was a Professor of computer science with the University of Leipzig, Germany, until 2002. Since 2002, he has been a Professor of computer science with the University of Konstanz, Germany. He is the coauthor of the book *Chaos and Fractals*, which won the Association of American Publishers Award for Best Mathematics Book of the Year, in 1992, and well over 100 research articles. His research interests include image and video processing, computer graphics, scientific visualization, dynamical systems, and sport informatics.



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 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and IEEE SIGNAL PROCESSING LETTERS.