RESEARCH ARTICLE

On the Distribution of Salient Objects in Web Images and Its Influence on Salient Object Detection

Boris Schauerte*, Rainer Stiefelhagen

Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Baden–Württemberg, Germany

* boris.schauerte@kit.edu

Abstract

In recent years it has become apparent that a Gaussian center bias can serve as an important prior for visual saliency detection, which has been demonstrated for predicting human eye fixations and salient object detection. Tseng et al. have shown that the photographer's tendency to place interesting objects in the center is a likely cause for the center bias of eye fixations. We investigate the influence of the photographer's center bias on salient object detection, extending our previous work. We show that the centroid locations of salient objects in photographs of Achanta and Liu's data set in fact correlate strongly with a Gaussian model. This is an important insight, because it provides an empirical motivation and justification for the integration of such a center bias in salient object detection algorithms and helps to understand why Gaussian models are so effective. To assess the influence of the center bias on salient object detection, we integrate an explicit Gaussian center bias model into two state-of-the-art salient object detection algorithms. This way, first, we quantify the influence of the Gaussian center bias on pixel- and segment-based salient object detection. Second, we improve the performance in terms of F_1 score, F_β score, area under the recall-precision curve, area under the receiver operating characteristic curve, and hit-rate on the well-known data set by Achanta and Liu. Third, by debiasing Cheng et al.'s region contrast model, we exemplarily demonstrate that implicit center biases are partially responsible for the outstanding performance of state-of-the-art algorithms. Last but not least, we introduce a non-biased salient object detection method, which is of interest for applications in which the image data is not likely to have a photographer's center bias (e.g., image data of surveillance cameras or autonomous robots).

1 Introduction

Among other influences such as task-specific factors, human attention is attracted to salient stimuli. In this context, saliency describes the subjective, perceptual quality that lets some items in the world stand out from their neighbors and immediately grab our attention. Accordingly,

(a) Example images

(b) Centroid scatter plot

(c) Example segmentation masks

(d) Mean segment mask
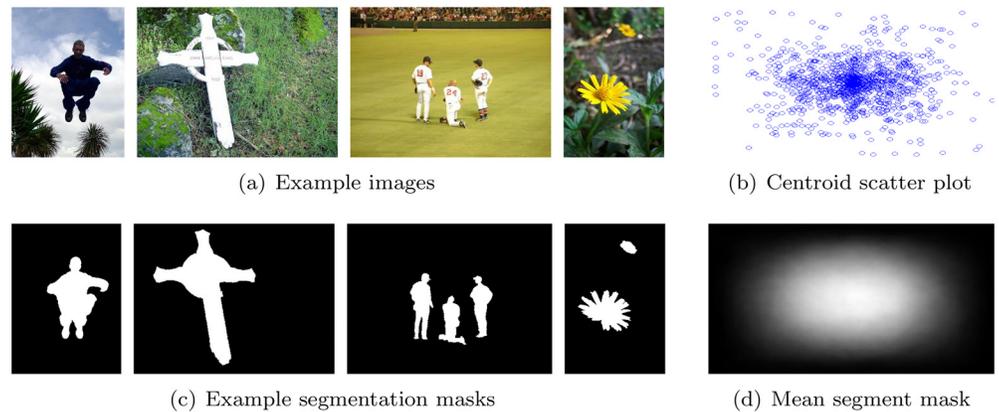
Fig 1. Illustration of the Achanta/Liu data set: example images (a), the corresponding segmentation masks (c), the mean over all segmentation masks (d), and the scatter plot of the centroid locations across all images (b).

doi:10.1371/journal.pone.0130316.g001

the goal of visual saliency detection is to determine what parts of an image are likely to grab the human attention. The task of "traditional" *visual saliency detection* is to predict where human observers look when presented with a scene, which can be recorded using eye tracking equipment (e.g., [1–4]). Liu et al. adapted the traditional definition of visual saliency by incorporating the high level concept of a salient object into the process of visual attention computation [5]. Here, a *salient object* is defined as being the object in an image that attracts most of the user's interest such as, for example, the man, the cross, the baseball players and the flowers in Fig 1. Accordingly, Liu et al. [5] defined the task of *salient object detection* as the binary labeling problem of separating the salient object from the background. Thus, in contrast to traditional visual saliency detection, salient object detection does not just comprise of the task to calculate the saliency of image regions, but it also incorporates the task to determine and segment the most salient object in the image. Here, it is important to note that the selection of a salient object happens consciously by the user whereas the gaze trajectories that are recorded using eye trackers are the result of mostly unconscious processes. Consequently, also taking into account that salient objects attract the human gaze (see, e.g., [1]), salient object detection and predicting where people look are very closely related yet substantially different tasks.

The photographer's center bias, i.e. the natural tendency of photographers to place the objects of interest near the center of their composition in order to enhance their focus and size relative to the background (see Tseng et al. [6]; we would like to note that Tseng et al.—due to their methodology—did not investigate the exact spatial distribution of the objects that attract the gaze, since they hired five persons who provided subjective scores from 1 to 5 in terms of how interesting things were biased toward the image center), has been identified as one cause for the often reported center bias in eye-tracking data during eye-gaze studies [7–9]. As a consequence, the integration of a center bias has become an increasingly important aspect in visual saliency models that focus on gaze prediction (e.g., [2, 3, 10]). In contrast, most recently proposed salient object detection algorithms do not incorporate an explicit model of the photographer's center bias (see, e.g., [11–14]). A notable exception and closely related to our work is the work by Jiang et al. [15], in which one of the three main criteria that characterize a salient object is that "it is most probably placed near the center of the image" [15]. The authors justify this characterization with the "rule of thirds", which is one of the most well-known principles of photographic composition (see, e.g., [16]), and use a Gaussian distance metric as a model. However, Jiang et al. do neither justify why the rule of thirds would be well represented by a

Gaussian distance metric nor do they investigate the quantitative influence of such a Gaussian center bias. We go beyond following the rule of third and show that the distribution of the objects' centroids correlates strongly positively with a 2-dimensional Gaussian distribution. This means nothing less than that we provide a strong empirical justification for integrating Gaussian center bias models into salient object detection algorithms. To demonstrate the importance, we adapt two state-of-the-art salient object detection methods to quantify the influence of the photographer's center bias on salient object detection.

The contribution of this paper is twofold: First, we use the salient object data set by Achanta et al. [11] to investigate the spatial distribution of salient objects in images. This way, in Sec. 3, we show that it is likely that salient objects in photographs are distributed around the image center in such a way that the radii are half-Gaussian distributed and the angles are uniformly distributed. Second, in Sec. 4, we explicitly integrate Gaussian center bias models in two recently proposed salient object detection methods: The pixel-based maximum symmetric sur-round salient object detection by Achanta et al. [12] and the segment-based region contrast method by Cheng et al. [14]. In order to measure the influence, we use the following evaluation measures: The maximum F_1 score, the maximum F_β score with $\beta = \sqrt{0.3}$ [11], the area-under-curve of the precision-recall curve, the AUC of the receiver operating characteristic (ROC AUC), and the hit-rate. In summary, the integration of the center bias model increases the ROC AUC by 2% and the performance with respect to all remaining measures by roughly 5%. Thus, we further advance the state-of-the-art of pixel-based as well as segment-based salient object detection. By modifying Cheng et al.'s region contrast model [14], first, we obtained a non-biased salient object detection algorithm that is based on region contrast and, second, we exemplarily demonstrate that implicit center biases can already be found in well-performing, state-of-the-art salient object detection algorithms and substantially influence the performance. This is important to consider when comparing and selecting algorithms for applications in which the data is not necessarily biased towards the center.

The remainder of this paper is organized as follows: In Sec. 2, we provide an overview of related work. Subsequently, in Sec. 3, we introduce and investigate our hypotheses about the spatial distribution of salient objects. Then, in Sec. 4, we integrate our hypotheses into two recently proposed salient object detection methods and evaluate the influence on the salient object detection performance. We conclude with a short summary and discussion in Sec. 5. Furthermore, please feel free to check the supplemental material for additional information such as, e.g., further evaluation results.

2 Related Work

We focus on the most recent related work that addresses bottom-up saliency detection with an emphasis on salient object detection (see, e.g., [17] for a more general overview of computa-tional attention models). Such methods may be biologically motivated, or purely computa-tional, or involve both aspects. In 2009, Achanta et al. [11, 12] introduced a salient object detection approach that basically relies on the difference of pixels to the average color and intensity value. In order to evaluate their approach, they selected a sub-set of 1000 images of the image data set that was collected from the web by Liu et al. [5] and calculated segmentation masks of the salient objects that were marked by 9 participants using (rough) rectangle annota-tions [5]. Please note that this procedure also means that during the manual data set annotation the selection of the salient object happens mostly conscious whereas gaze trajectories that are recorded using eye trackers are a result of a mostly unconscious process. Since it was created, the salient object data set by Achanta et al. serves as reference data set to evaluate methods for salient object detection (see, e.g., [11–14]). Liu et al. [5] and Alexe et al. [18] approach salient

object detection using machine learning. To this end, Liu et al. [5] combine multi-scale contrast, center-surround histograms, and color spatial-distributions with conditional random fields. Similarly, Alexe et al. [18] combine multi-scale saliency, color contrast, edge density, and superpixels in a Bayesian framework. Closely related to Bayesian surprise [19], Klein et al. [13] use the Kullback-Leibler Divergence of the center and surround image patch histograms to calculate the saliency. Cheng et al. [14] use segmentation to define a regional contrast-based method, which simultaneously evaluates global contrast differences and spatial coherence. Here, we can differentiate between algorithms that rely on segmentation-based (e.g., [14, 18]) and pixel-based contrast measures (e.g., [11–13]). Closely related to our work on the quantitative influence of the center bias on salient object detection is the work by Jiang et al. [15] and most recently Borji et al. [20]. In Jiang et al.'s work [15] one of the main criteria that characterize a salient object is that "it is most probably placed near the center of the image", which is justified with the "rule of thirds". Most recently, Borji et al. [20] evaluated several salient object detection models and also performed tests with an additive Gaussian center bias and conclude that the resulting "change in accuracy is not significant and does not alter model rankings". But, this neglects the possibility that well-performing models already have an integrated, implicit center bias, which—as one part of our work—we demonstrate exemplarily to be the case for Cheng et al.'s region contrast algorithm [14]. Furthermore, there exist several approaches that explicitly integrate a center bias, but do not provide a quantitative evaluation of its influence nor an empirical justification of the chosen model (e.g., [21]). In this paper, we adapt the pixel-based method by Achanta et al. [12] and the segmentation-based method by Cheng et al. [14] to incorporate a model of the photographer-related center bias and quantify the influence of the center bias on the performance. Furthermore, Borji et al. [20] do not provide an empirical justification why a Gaussian distribution is an appropriate center bias model, which is another part of the work described in this paper.

It has been observed in several studies that the visual attention of human participants in natural scenes is biased toward the center of static images and videos (see, e.g., [8, 9, 22]). One possible bottom-up cause of the bias is intrinsic bottom-up visual saliency as predicted by computational saliency models. One possible top-down cause of the center bias is known as photographer bias (see, e.g., [7–9]), which describes the natural tendency of photographers to place objects of interest near the center of their composition. In fact, what the photographer considers interesting may also be highly bottom-up salient. Additionally, the photographer bias may lead to a viewing strategy bias [23], which means that viewers may orient their attention more often toward the center of the scene, because they expect salient or interesting objects to be placed there. Thus, since in natural images and videos the distribution of objects of interest and thus saliency is usually biased toward the center, it is often unclear how much the saliency actually contributes in guiding attention. It is possible that people look at the center for reasons other than saliency, but their gaze happens to fall on salient locations. Therefore, this center bias may result in overestimating the influence of saliency computed by the model and contaminate the evaluation of how visual saliency may guide orienting behavior. Recently, Tseng et al. [6] were able to demonstrate quantitatively that center bias is correlated strongly with photographer bias and is influenced by viewing strategy at scene onset. Furthermore, e.g., they were able to show that motor bias had no effect. However, they did not evaluate and computationally model how specifically the objects that attract the gaze are distributed spatially in the image. Instead, Tseng et al. hired five naive participants to provide subjective scores from 1 to 5 in terms of how interesting things were biased toward the image center [6]. In this paper, we use the data set by Achanta et al. [11] to investigate the distribution of salient objects in photographs and then evaluate the influence on two state-of-the-art salient object detection models.

3 Center Bias Model

To investigate the spatial distribution of salient objects in photographs collected from the web, we use the manually annotated segmentation masks by Achanta et al. [11, 12] that mark the salient objects in 1000 images of the salient object data set by Liu et al. [5]. More specifically, we use the segmentation masks to determine the centroids of all salient objects in data set and analyze the centroids' spatial distribution. The images in the data set by Liu et al. [5] have been collected from a variety of sources, mostly from image forums and image search engines. Liu et al. collected more than 60,000 images and subsequently selected an image subset in which all images contain a salient object or a distinctive foreground object [5]. Nine users marked the salient objects using (rough) bounding boxes and the salient objects in the image database have been defined based on the "majority agreement". However, as a consequence of the selection process, the data set does not include images without distinct salient objects. This is an important aspect to consider when trying to generalize the results reported on Achanta et al.'s and Liu et al.'s data set to other data sets or application areas.

In order to statistically analyze the 2-dimensional spatial distribution of the salient objects' centroids, we first identify the center of the spatial distribution. Then, given the distribution's center, we can use a polar coordinate system to independently analyze the distribution of the angles and distances between the center and the salient objects.

3.1 The Center

Our model is based on a polar coordinate system that has its pole at the image center. Since the images in Achanta's data set have varying widths and heights, we use in the following normalized Cartesian image coordinates in the range $[0, 1] \times [0, 1]$. The mean salient object centroid location is $[0.5021, 0.5024]^T$ and the corresponding covariance matrix is $[0.0223, -0.0008; -0.0008, 0.0214]$. Thus, we can motivate the use of a polar coordinate system that has its pole at $[0.5, 0.5]^T$ to represent all locations relative to the expected distribution's mode.

3.2 The Angles are Distributed Uniformly

Our first model hypothesis is that the centroids' angles in the specified polar coordinate system are uniformly distributed in $[-\pi, \pi]$.

In order to investigate the hypothesis, we use a Quantile-Quantile (Q-Q) plot as a graphical method to compare probability distributions (see [24]). In Q-Q plots the quantiles of the samples of two distributions are plotted against each other. Thus, the more similar the two distributions are, the better the points in the Q-Q plot will approximate the line $f(x) = x$. We calculate the Q-Q plot of the salient object location angles in our polar coordinate system versus uniformly drawn samples in $[-\pi, \pi]$, see Fig 2 (left). The apparent linearity of the plotted Q-Q points supports the hypothesis that the angles are distributed uniformly.

We can quantify the observed linearity, see Fig 2 (left), to analyze the correlation between the model distribution and the data samples using probability plot correlation coefficients (PPCC) [24]. The PPCC is the correlation coefficient between the paired quantiles and measures the agreement of the fitted distribution with the observed data (i.e., goodness-of-fit). The closer the correlation coefficient is to one, the higher the positive correlation and the more likely the distributions are shifted and/or scaled versions of each other. Furthermore, by comparing against critical values of the PPCC (see [25] and [24]), we can use the PPCC as a statistical test, which is closely related to the Shapiro-Wilk test [26] and can reject the hypothesis that the data samples match the assumed model distribution. Furthermore, we can use the correlation to test the hypothesis of no correlation by transforming the correlation to create a t-statistic.
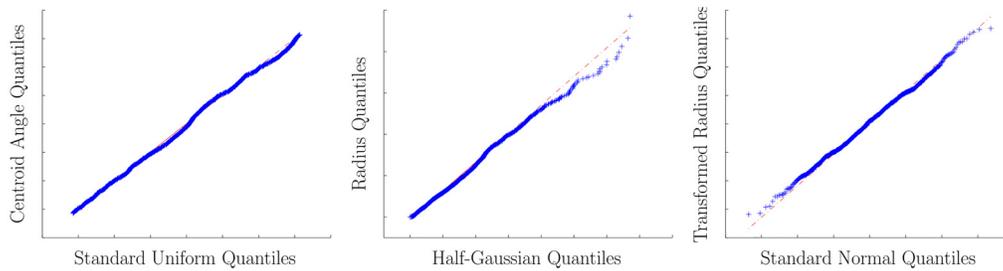
Fig 2. Quantile-Quantile (Q-Q) plots of the angles versus a uniform distribution (left), radii versus a half-Gaussian distribution (middle), transformed radii (see Sec. 3.3) versus a normal distribution (right).

doi:10.1371/journal.pone.0130316.g002

The obvious linearity of the Q-Q plot, see Fig 2 (left), is reflected by a PPCC of 0.9988 (mean of several runs with $N = 1000$ uniform randomly selected samples), which is substantially higher than the critical value of 0.8880 (see [25]) and thus the hypothesis of identical distributions can not be rejected. Furthermore, the hypothesis of no correlation is rejected at $\alpha = 0.05$ ($p = 0$).

3.3 The Radii follow a Half-Gaussian Distribution

Our second model hypothesis is that the radii of the salient object locations follow a half-Gaussian distribution. We have to consider a half-Gaussian distribution in the interval $[0, \infty]$, because the radius—as a length—is by definition positive. If we consider the image borders, we could assume a two-sided truncated distribution, but we have three reasons to work with a one-sided model: The variance of the radii seems sufficiently small, the "true" centroid of the salient object may be outside the image borders (i.e., parts of the salient object can be truncated by the image borders), and it facilitates the use of various, well-known statistical tests (see [27]).

We can use a Q-Q plot against a half-Gaussian distribution to graphically assess the hypothesis, see Fig 2 (middle). The linearity of the points suggests that the radii are distributed according to a half-Gaussian distribution. The visible outliers in the upper-right are caused by less than 30 centroids that are highly likely to be disturbed by the image borders. Please be aware of the fact that it is not necessary to know the exact distribution parameters when working with Q-Q plots as long as the distributions are linearly related (see [24]). Furthermore, we transform the polar coordinates in such a way that they represent the same point with a combination of positive angles in $[0, \pi]$ and radii in $[-\infty, \infty]$. This way, we can compare the distribution of the transformed radii against a normal distribution with its mode and mean at 0, see Fig 2 (right).

The obvious correlation that is visible in the Q-Q plots, see Fig 2 (middle and right), is reflected by a PPCC of 0.9987, which is above the critical value of 0.9984 (see [24]). The hypothesis of no correlation is rejected at $\alpha = 0.05$ ($p = 0$).

4 Quantifying the Influence on Salient Object Detection

To assess the influence of the center bias on pixel- and object-based salient object detection, we integrate a Gaussian center bias into the algorithms by Achanta et al. [12] and Cheng at al. [14].

4.1 Center Biased Saliency Models

Pixel-based. As a pixel-based model, we use maximum symmetric surround saliency detection by Achanta et al. [12] in combination with a Gaussian center bias map (cf., e.g., [3, 10]). To this end, we define the center bias saliency map $S_C \in \mathbf{R}^{M \times N}$

$$S_C(x, y) = g(\mu_x - x, \mu_y - y; \sigma_x, \sigma_y) \quad \text{with} \tag{1}$$

$$g(x, y; \sigma_x, \sigma_y) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left\{-\frac{1}{2}\frac{x^2}{\sigma_x^2}\right\}$$
$$* \frac{1}{\sqrt{2\pi}\sigma_y}\exp\left\{-\frac{1}{2}\frac{y^2}{\sigma_y^2}\right\} \qquad (2)$$

where (x, y) is the pixel coordinate, $\mu = (\mu_x, \mu_y)$ is the image center's coordinate, and σ_x and σ_y are the standard deviation in x- and y-direction depending on the image width and height, respectively.

In order to investigate the influence of the center bias, we investigate different, plausible strategies to investigate the combination of the bottom-up and center bias saliency maps S_B and S_C, respectively:

$$S_P = f(S_C, S_B) \qquad (3)$$

where f is the chosen center bias integration scheme.

We consider the following schemes, cf. [4]: First, a convex, linear integration, i.e. $f_+(S_C, S_B) = w_C S_C + w_B S_B$ with $w_B + w_C = 1$ ($w_B, w_C \in \mathbf{R}_0^+$). Second, multiplicative integration as a supra-linear combination method, i.e. $f_\circ(S_C, S_B) = S_C \circ S_B$, where $\circ$ denotes the Hadamard product. Third, the minimum as a further, alternative supra-linear combination, i.e. $f_\downarrow(S_C, S_B) = \min(S_C, S_B)$. Fourth, the maximum to realize a late, sub-linear combination scheme, i.e. $f_\uparrow(S_C, S_B) = \max(S_C, S_B)$. All these schemes are also related to different Fuzzy logic interpretations, which might provide a common theoretical framework and interpretation throughout later applications (e.g., [28]). To improve the readability, we refer to the linear combination for explicit center bias integration—unless stated otherwise, of course—in the following.

Segmentation-based. As a segmentation-based model, we adapt Cheng et al.'s region contrast model [14]. This model is particularly interesting, because it already provides state-of-the-art performance, which is partially caused by an implicit (i.e., unmotivated, undiscussed and potentially unknowingly introduced by the authors) center bias as we will show in the following. This way, we can observe how the model behaves if we remove the implicit center bias—which was neither motivated nor explained by the authors—and add an explicit Gaussian center bias. The spatially weighted region contrast saliency equation is defined as follows

$$S_S(r_k) = \sum_{r_k \neq r_i} \hat{D}_s(r_k; r_i) w(r_i) D_r(r_k; r_i) \qquad \text{with} \qquad (4)$$

$$\hat{D}_s(r_k; r_i) = \exp\left(-D_s(r_k; r_i)/\sigma_s^2\right) \qquad (5)$$

$w(r_i)$ is the weight of region r_i, which equals the number of pixels in r_i—i.e., $w(r_i) = |r_i|$—to emphasize color contrast to bigger regions. $D_r(\cdot; \cdot)$ is the color distance metric between the two regions

$$D_r(r_1; r_2) = \sum_{c_1} \sum_{c_2} f(c_{1;i}) f(c_{2;j}) D(c_i; c_j) \qquad (6)$$

where $f(c_{k;i})$ is the (frequentist) probability of the i-th color $c_{k;i}$ among all n_k colors in the k-th region r_k, which is determined using a color histogram. The probability of the color inside the regions $f(c_{k;i})$ is used as weight to emphasize color differences between dominant colors. $D(c_i; c_j)$ measures the distance between the colors and in the following it is defined as being the Euclidean distance in the CIE Lab color space. Finally, $\hat{D}_s(r_k; r_i)$ is the spatial distance between regions r_k and r_i, where σ_s controls the spatial weighting. The spatial distance between two regions is defined as the Euclidean distance between the centroids of the respective regions
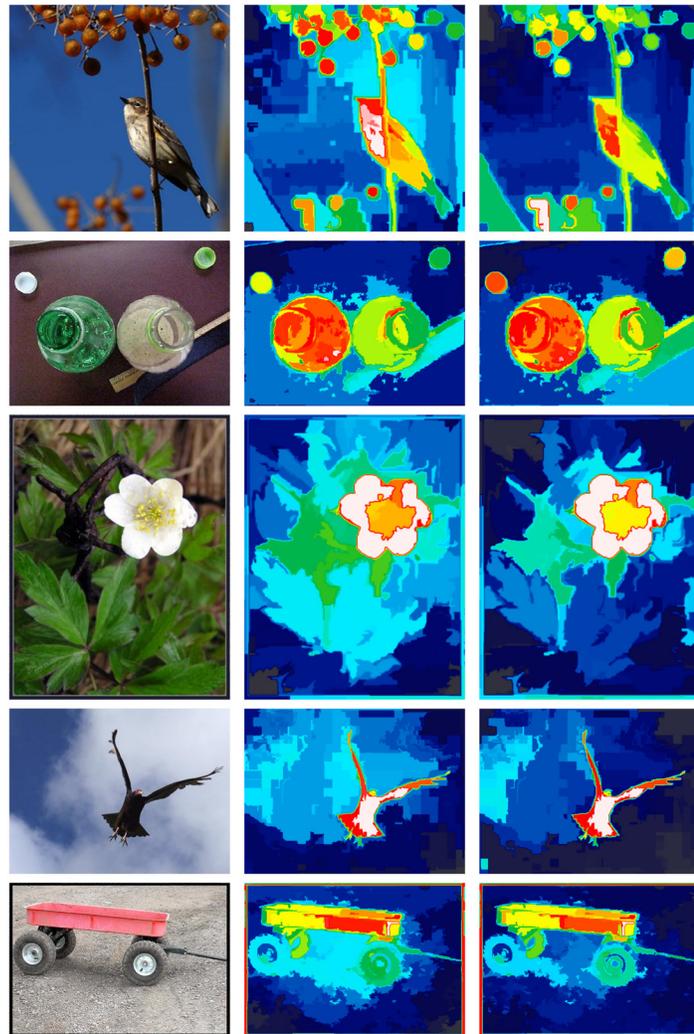
Fig 3. An example illustrating the influence of the implicit center bias in the region contrast method by Cheng et al. [14]. Left-to-right: Image, region contrast (RC) saliency map, and locally debiased region contrast (LDRC) saliency map. As can be seen, RC tends to assign a comparatively high saliency to regions in the center of the image even if these regions exhibit an apparently low perceptual saliency such as, for example, the space between the hand cart's wheels in the bottom row.

doi:10.1371/journal.pone.0130316.g003

using pixel coordinates that are normalized to the range $[0, 1] \times [0, 1]$. Smaller values of σ_s influence the spatial weighting in such a way that the contrast to regions that are farther away contributes less to the saliency of the current region.

It is this unnormalized Gaussian weighted Euclidean distance $\hat{D}_s(r_k; r_i)$ that causes an implicit Gaussian-like center bias (see Figs 3 and 4), because it favors regions whose distances to the other neighbors are smaller, which is—in general—the case for segments at the center of the image. Although this biased distance function has a significant impact on the performance, its choice has not been clearly motivated, discussed, or evaluated by Cheng et al. To remove this implicit bias, we introduce a normalized, i.e. locally debiased, distance function $\check{D}_s(r_k; r_i)$ that still weights close-by regions higher than further away regions, but does not lead to an
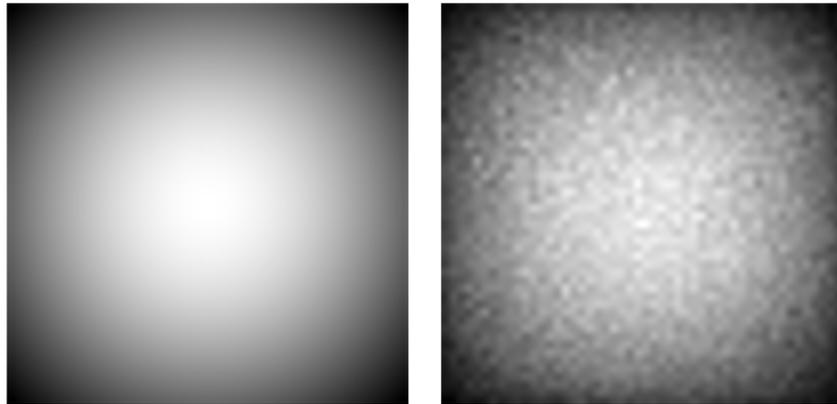
Fig 4. Illustration of the implicit center bias in the method by Cheng et al. [14]. Left: Each pixel shows the distance weight sum, i.e. $\sum_{r_i} \hat{D}_s(r_k; r_i)$, to all other pixels in a regular grid. Right: The average weight sum depending on the centroid location calculated on the Achanta/Liu data set using Felzenszwalb's segmentation method.

doi:10.1371/journal.pone.0130316.g004

implicit center bias

$$\check{D}_s(r_k; r_i) = \frac{\hat{D}_s(r_k; r_i)}{\sum_{r_i} \hat{D}_s(r_k; r_i)} \qquad (7)$$

i.e.
$$\forall r_k : \sum_{r_i} \check{D}_s(r_k; r_i) = 1 \qquad (8)$$

Similar to the pixel-based model (see Sec. 4.1), we can now integrate an explicit center bias into the segmentation-based model

$$S_S(r_k) = f\left(\sum_{r_k \neq r_i} \check{D}_s(r_k; r_i) w(r_i) D_r(r_k; r_i), \ g(C(r_k); \sigma_x, \sigma_y) \right) \qquad (9)$$

Here, f is the chosen center bias integration function as in Eq 3. Furthermore, $C(r_k)$ denotes the centroid of region r_k and g is defined as in Eq 2.

4.2 Evaluation Procedure

Dataset. As for the graphical investigation of our hypotheses using Q-Q plots (see Fig 2), we use the manually annotated segmentation masks by Achanta et al. [11, 12], see Sec. 3, to quantify the influence of the Gaussian center bias on salient object detection.

Baseline algorithms. In order to compare our results, we use a set of saliency detection algorithms that we group into two coarse categories: First, algorithms that were specifically proposed for salient object detection and, second, algorithms that have been proposed and evaluated in other contexts. From the second category, we use: The well-known saliency model by Itti and Koch [29], Graph-Based Visual Saliency (GBVS) by Harel at al. [30], Context-Aware Saliency (CAS) by Goferman et al. [31, 32], and the FFT's spectral residuals (FFT) and DCT image signatures (DCT) by Hou et al. [33, 34]. For FFT and DCT, we optimized the resolution at which the saliency maps are calculated, which is the most important algorithm parameter and has a significant influence on the performance. As baseline for salient object detection algorithms (first category), we use: The Frequency-Tuned model (FT) by Achanta et al. [11] (please

note that an erratum regarding their reported results has been published at http://ivrg.epfl.ch/supplementary_material/RK_CVPR09), the Bonn Information-Theoretic Saliency model (BITS) by Klein et al. [13], the Maximum Symmetric Surround Saliency (MSSS) model by Achanta et al. [12], and the Region Contrast (RC) model by Cheng et al. [14] that uses Felzenszwalb's image segmentation method [35]. The latter two are the original algorithms we adapted.

Of course, we evaluate our adapted, center biased models: The maximum symmetric surround saliency with center bias (MSSS+CB; see Sec. 4.1) and the region contrast model with explicit center bias (RC+CB; see Sec. 4.1). In order to investigate the influence of the implicit center bias in the region contrast model (see Sec. 4.1), we calculate the performance of the locally debiased region contrast model without and with explicit center bias (LDRC and LDRC+CB, respectively; see Sec. 4.1). Additionally, as a reference we provide the results for the standalone segment-based and pixel-based center bias models, i.e. $w_C = 1$ (CB$_S$ and CB$_P$, respectively).

If available, we used the reference implementations that have been provided by the authors. For MSSS we use the C++ implementation by Achanta, because it provides a better performance than the basic Matlab implementation. For Itti we use the iLab Neuromorphic Vision Toolkit (iNVT). We integrated the methods directly into Matlab (mex) in order to avoid quantization and/or compression artifacts that may occur due to saving and loading them as images. For DCT and FFT, we used the implementations in our publicly available Matlab toolbox [36]. All calculations have been made using double precision arithmetic. To make our results as reproducible as possible (we have observed that the precision-recall curves of different authors vary), we will make our implementations and evaluation scripts open source. We would like to note that our evaluation measure implementations follow the implementations of Weka and LingPipe. The corresponding precision-recall curves and results of further baseline algorithms can be seen in Fig 5.

Measures. We can use the binary segmentation masks for saliency evaluation by treating the saliency maps as binary classifiers. At a specific threshold t we regard all pixels that have a saliency value above the thresholds as positives and all pixels with values below the thresholds as negatives. By sweeping over all thresholds $\min(S) \leq t \leq \max(S)$, we can evaluate the performance using common binary classifier evaluation measures.

Most commonly, precision-recall curves are used—e.g., by Achanta et al. [11, 12], Cheng et al. [14], and Klein et al. [13]—to evaluate the salient object detection performance. We use five evaluation measures to quantify the performance of the algorithms. We calculate the area under curve (AUC) of the (interpolated) precision-recall curve (PR) and the receiver operating characteristic (ROC) curve [37]. Complementary to the PR AUC, we calculate the maximum F_1 and $F_{\sqrt{0.3}}$ scores with

$$F_\beta = (1 + \beta^2) \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \tag{10}$$

F_β with $\beta = \sqrt{0.3}$ has been proposed by Achanta et al. to weight precision more than recall for salient object detection [11]. Additionally, we calculate the hit-rate (HR) that measures how often the pixel with the maximum saliency belongs to the salient object.

4.3 Quantitative Evaluation Results and Discussion

Explicit center bias integration type. How does the performance depend on the chosen center bias integration? To investigate this question, we tested the minimum, maximum, and product as alternative combinations. To account for the influence of different value distributions within the normalized value range, we also weighted the input of the min and max operation (e.g., $S_P^{\min} = \min(w_C S_C, w_B S_B)$). The results of the algorithms using different combination
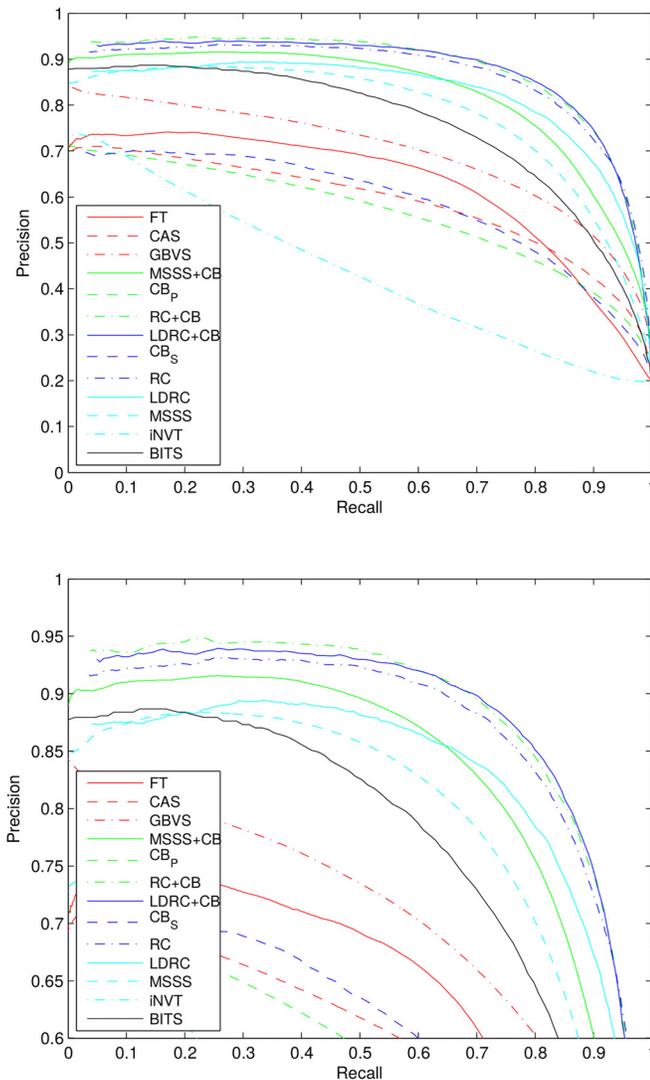
Fig 5. Precision-recall curves for all evaluated models with full (top) and limited range of the precision (bottom). This graphic is best viewed in color.

types are shown in Table 1. The presented results are the results that we achieve with the center bias weight that results in the highest F_1 score.

In Table 1, we can see that the linear combination is the best choice for LDRC+CB. However, for MSSS+CB and RC+CB the product seems to be the combination that provides the best performance. Apparently MSSS+CB benefits more from using the product as combination type than RC+CB. Also interesting to note is that LDRC+CB with the product as combination achieves similar results to RC. However, LDRC+CB remains the algorithm that provides the best performance in terms of F_1 score and F_β score whereas RC+CB provides the best performance in terms of PR AUC and HR. Interestingly, LDRC+CB and RC+CB achieve a nearly identical ROC AUC.

Convex center bias weight. How does the weight of the center bias influence the performance? To answer this question, we calculated the performance of LDRC+CB, RC+RB, and

Table 1. The maximum F_1 score, maximum F_β score, PR AUC ($\int$PR), ROC AUC ($\int$ROC), and Hit-Rate (HR) that we obtain using different combination types.

Method	Combination	F_1	F_β	$\int$PR	$\int$ROC	HR
LDRC+CB	Linear/Convex	0.8034	0.8183	0.8800	0.9624	0.9240
LDRC+CB	Max	0.7504	0.7561	0.8108	0.9422	0.8630
LDRC+CB	Min	0.7897	0.8049	0.8584	0.9535	0.8880
LDRC+CB	Product	0.7883	0.8024	0.8704	0.9578	0.9130
LDRC	–	0.7574	0.7675	0.8302	0.9430	0.8680
RC+CB	Linear/Convex	0.7973	0.8120	0.8833	0.9620	0.9340
RC+CB	Max	0.7855	0.7993	0.8710	0.9568	0.9140
RC+CB	Min	0.7962	0.8150	0.8807	0.9603	0.9180
RC+CB	Product	0.7974	0.8136	0.8878	0.9623	0.9460
RC	–	0.7855	0.7993	0.8710	0.9568	0.9140
MSSS+CB	Linear/Convex	0.7490	0.7678	0.8265	0.9495	0.8900
MSSS+CB	Max	0.7165	0.7337	0.7849	0.9270	0.8420
MSSS+CB	Min	0.7373	0.7606	0.8211	0.9339	0.9140
MSSS+CB	Product	0.7523	0.7748	0.8398	0.9445	0.9350
MSSS	–	0.7165	0.7337	0.7849	0.9270	0.8420
CB_S	–	0.5793	0.5764	0.5920	0.8623	0.6980
CB_P	–	0.5604	0.5452	0.5638	0.8673	0.7120

doi:10.1371/journal.pone.0130316.t001

MSSS+CB with $w_C \in [0, 1]$ in 0.025 steps. The resulting curves of the F_1 score, F_β score, PR AUC, ROC AUC, and hit-rate are shown in Fig 6(c), 6(a) and 6(b), respectively.

For each of the three algorithms the values of w_C that lead to the optimal F_1 score, F_β score, PR AUC, and ROC AUC lie within a small interval. In contrast, for all algorithms the value of w_C that achieves the highest hit-rate is outside these intervals and substantially higher. Furthermore, the best weight for each measure depends on the algorithm and varies substantially. It is interesting to see that small weights only have a minor (yet positive) influence on RC+CB until a point is reached (roughly at $w_C = 0.55$) where the performance begins to drop significantly. This becomes especially apparent when comparing the curves of RC+CB, see Fig 6(a), with the curves of LDRC+CB, see Fig 6(c).

Quantitative comparison. The center bias itself already has a considerable predictive power, see Table 2, and is relatively close to the performance of FT. However, there is a substantial performance gap between the standalone center bias models (CB_S and CB_P) and good non-biased methods such as, e.g., MSSS and LDRC.

As could be expected, the performance of RC drops substantially if we remove the implicit center bias as is done by LDRC (see Sec. 4.1), which can best be seen in Table 3. What happens if we add our explicit center bias model to unbiased models? As can be seen in the performance difference between MSSS and MSSS+CB as well as the performance difference between LDRC to LDRC+CB, the performance is substantially increased with respect to all evaluation measures, see Tables 2 and 3. Interestingly, the relative performance improvement from pixel-based MSSS to MSSS+CB and segment-based LDRC to LDRC+CB is comparable, see Table 3. Furthermore, with the exception of HR, the performance of LDRC+CB and RC+CB is nearly identical with a slight advantage for LDRC+CB (see Tables 2 and 3). This indicates that we did not lose important information by debiasing the distance metric (LDRC+CB vs RC+CB) and that the explicit Gaussian center bias model is advantageous compared to the implicit weight bias (LDRC+CB and RC+CB vs RC).
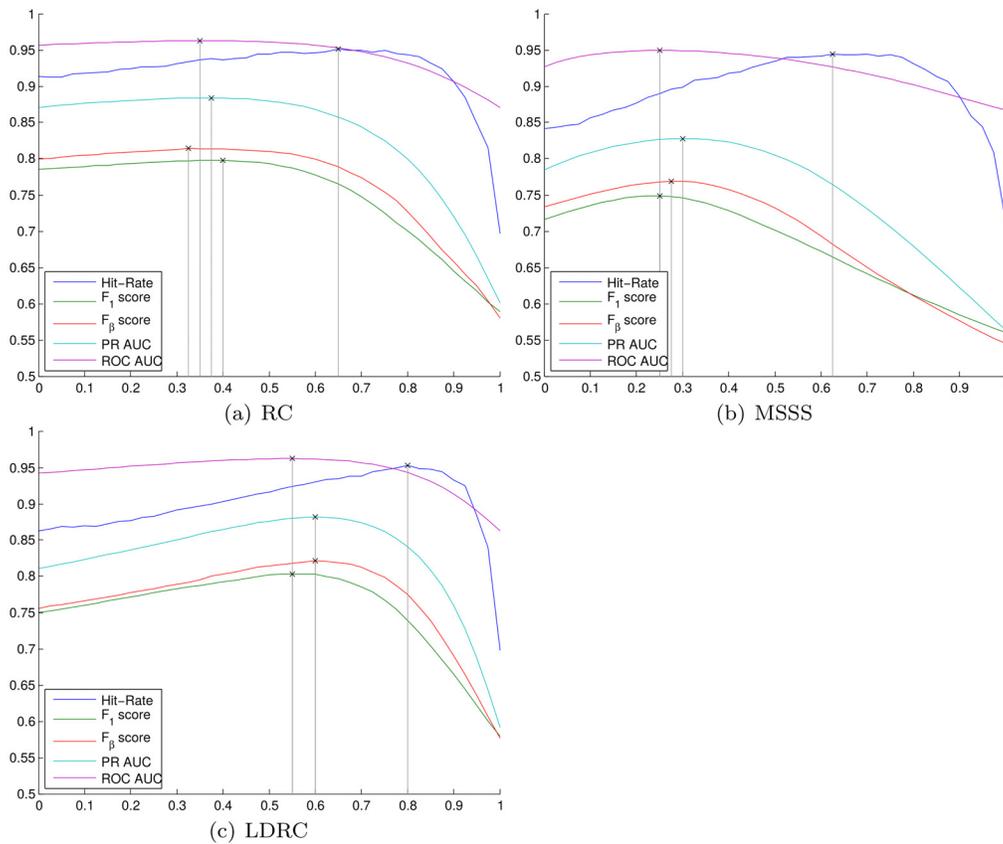
Fig 6. Illustration of the influence of the weight w_C on the performance of RC+CB, LDRC+CB, and MSSS+CB (convex combination).

doi:10.1371/journal.pone.0130316.g006

In summary, MSSS+CB provides a substantially higher performance than MSSS and outperforms, e.g., FT and BITS. RC+CB and LDRC+CB provide a better performance than their unbiased counterparts RC and LDRC, respectively. Furthermore, their performance is very similar and both outperform all other models. Interestingly, LDRC is the best model without center bias in our evaluation on Achanta's data set. This makes LDRC an interesting candidate for applications in which the image data can not be expected to have a photographer's center bias (e.g., image data of surveillance cameras, autonomous robots, or human-robot interaction [38]).

Statistical significance. One question remains: Does the integration of an explicit center bias result in a statistically significant performance improvement? To address this question, we test the performance (i.e., F_1, F_β, $\int$PR, and $\int$ROC) of LDRC and MSSS with and without an explicit center bias. For this purpose, we rely on two pairwise, two-sample t-tests: First, we perform a two-tailed test to check whether the compared performances with and without an integrated center bias come from distributions with equal means (i.e., $\mathcal{H}_=$: "means are equal"). Second, we perform a one-tailed test to check whether the performance with an integrated center bias is worse than without an integrated center bias, i.e. the center biased performance distribution's mode is lower (i.e., $\mathcal{H}_<$: "mean is lower"). If we can reject both hypotheses, then it is clear that the performance of the algorithm has significantly improved due to the integrated center bias. All tests are performed at a confidence level of 95%, i.e., $\alpha = 5\%$.

For MSSS, we can reject the hypothesis of equal mean for F_1, F_β, $\int$PR, and $\int$ROC with $p_{F_1} = 0.0285$, $p_{F_\beta} = 0.0031$, $p_{\int\text{PR}} = 5.252 \times 10^{-7}$, and $p_{\int\text{ROC}} = 2.618 \times 10^{-16}$, respectively. Additionally,

Table 2. The maximum F_1 score, maximum F_β score, PR AUC ($\int$PR), ROC AUC ($\int$ROC), and Hit-Rate (HR) of the evaluated algorithms (sorted ascending by F_β).

Method	F_1	F_β	$\int$PR	$\int$ROC	HR
LDRC+CB	0.8034	0.8183	0.8800	0.9624	0.9240
RC+CB	0.7973	0.8120	0.8833	0.9620	0.9340
RC	0.7855	0.7993	0.8710	0.9568	0.9140
MSSS+CB	0.7490	0.7678	0.8265	0.9495	0.8900
LDRC	0.7574	0.7675	0.8302	0.9430	0.8680
BITS	0.7342	0.7582	0.7589	0.9316	0.7540
MSSS	0.7165	0.7337	0.7849	0.9270	0.8420
FFT	0.6455	0.6375	0.6593	0.8926	0.8080
DCT	0.6472	0.6368	0.6612	0.8962	0.8270
GBVS	0.6403	0.6242	0.6970	0.9088	0.8480
FT	0.5995	0.6009	0.6261	0.8392	0.7100
CB_S	0.5793	0.5764	0.5920	0.8623	0.6980
CAS	0.5857	0.5615	0.5888	0.8741	0.6920
CB_P	0.5604	0.5452	0.5638	0.8673	0.7120
iNVT	0.3383	0.4012	0.4396	0.5768	0.6870

doi:10.1371/journal.pone.0130316.t002

we can reject the hypothesis that an integrated center bias has a negative influence on the performance with $p_{F_1} = 0.0142$, $p_{F_\beta} = 0.0015$, $p_{\int PR} = 2.626 \times 10^{-7}$, and $p_{\int ROC} = 1.309 \times 10^{-16}$.

Similarly, we can reject the hypothesis that the performance of LDRC with and without center bias has an equal mean for F_1, F_β, $\int PR$, and $\int ROC$ with $p_{F_1} = 0.0018$, $p_{F_\beta} = 2.426 \times 10^{-5}$, $p_{\int PR} = 1.118 \times 10^{-7}$, and $p_{\int ROC} = 1.555 \times 10^{-5}$, respectively. And, we can reject the hypothesis that an integrated center bias has a negative influence on the performance with $p_{F_1} = 9.071 \times 10^{-4}$, $p_{F_\beta} = 1.213 \times 10^{-5}$, $p_{\int PR} = 5.590 \times 10^{-8}$, and $p_{\int ROC} = 7.773 \times 10^{-6}$.

Consequently, it is apparent that the integration of a center bias can lead to statistically significant performance improvements for pixel-based as well as segmentation-based algorithms.

5 Conclusion

We formulated and investigated two hypotheses about the location of salient objects in photographs: First, the radial centroid distribution around the image center is uniform. Second, the distances between their centroids and the image center follow a normal distribution. We investigated these hypotheses using graphical methods, which indicate that our hypotheses are true. This is an important insight, because it provides a strong empirical motivation and justification for the widely applied Gaussian center bias models. To investigate the influence of the center bias on salient object detection, we explicitly integrated the center bias model in two state-of-the-art salient object detection algorithms. We have shown that the explicitly modeled center

Table 3. Relative performance (in %) of our adapted algorithms with respect to their baseline.

Method	Baseline	F_1	F_β	$\int$PR	$\int$ROC	HR
LDRC	RC	96.4	96.0	95.3	98.6	95.0
RC+CB	RC	101.5	101.6	101.4	100.5	102.2
LDRC+CB	RC	102.3	102.4	101.0	100.6	101.1
LDRC+CB	LDRC	106.1	106.6	106.0	102.1	106.5
MSSS+CB	MSSS	104.5	104.7	105.3	102.4	105.7

doi:10.1371/journal.pone.0130316.t003

bias has a significant, positive influence on the performance (in terms of hit-rate, the area under the precision-recall curve, the area under the receiver operating characteristic curve, the F_1 score, and the F_β score). Last but not least, by debiasing Cheng et al.'s region contrast model, we have exemplarily shown that implicit center biases might at least partially be responsible for the performance of state-of-the-art salient object detection algorithms and as a consequence we introduced an adapted, non-biased salient object detection algorithm.

Acknowledgments

Author Contributions

References

1. Einhäuser W, Spain M, Perona P. Objects predict fixations better than early saliency. Journal of Vision. 2008; 8(14).

2. Yang, Y, Song M, Li N, Bu J, Chen C. What is the chance of happening: a new way to predict where people look. In: Proc. European Conf. Comp. Vis.; 2010.

3. Judd T, Ehinger K, Durand F, Torralba A. Learning to Predict Where Humans Look. In: Proc. Int. Conf. Comp. Vis.; 2009.

4. Schauerte B, Stiefelhagen R. Predicting Human Gaze using Quaternion DCT Image Signature Saliency and Face Detection. In: Proc. Workshop on the Applications of Computer Vision; 2012.

5. Liu T, Sun J, et al. Learning to Detect A Salient Object. In: Proc. Int. Conf. Comp. Vis. Pat. Rec.; 2007.

6. Tseng PH, Carmi R, Cameron IGM, Munoz DP, Itti L. Quantifying center bias of observers in free viewing of dynamic natural scenes. Journal of Vision. 2009; 9(7). doi: 10.1167/9.7.4 PMID: 19761319

7. Reinagel P, Zador AM. Natural Scene Statistics At the Centre of Gaze. In: Network: Computation in Neural Systems; 1999. p. 341–350.

8. Parkhurst DJ, Niebur E. Scene content selected by active vision. Spatial Vision. 2003; 16(2):125–154. doi: 10.1163/15685680360511645 PMID: 12696858

9. Tatler BW. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision. 2007; 7(14). Available from: http://www.journalofvision.org/content/7/14/4.abstract. doi: 10.1167/7.14.4 PMID: 18217799

10. Borji A, Sihite DN, Itti L. Probabilistic Learning of Task-Specific Visual Attention. In: Proc. Int. Conf. Comp. Vis. Pat. Rec.; 2012.

11. Achanta R, Hemami S, Estrada F, Süsstrunk S. Frequency-tuned Salient Region Detection. In: Proc. Int. Conf. Comp. Vis. Pat. Rec.; 2009.

12. Achanta R, Süsstrunk S. Saliency detection using maximum symmetric surround. In: Proc. Int. Conf. Image Process.; 2010.

13. Klein DA, Frintrop S. Center-surround Divergence of Feature Statistics for Salient Object Detection. In: Proc. Int. Conf. Comp. Vis.; 2011.

14. Cheng MM, Zhang GX, Mitra NJ, Huang X, Hu SM. Global Contrast based Salient Region Detection. In: Proc. Int. Conf. Comp. Vis. Pat. Rec.; 2011.

15. Jiang H, Wang J, Yuan Z, Liu T, Zheng N. Automatic salient object segmentation based on context and shape prior. In: Proc. British Mach. Vis. Conf.; 2011.

16. Luo Y, Tang X. Photo and Video Quality Evaluation: Focusing on the Subject. In: Proc. European Conf. Comp. Vis.; 2008.

17. Tsotsos JK. A Computational Perspective on Visual Attention. The MIT Press; 2011.

18. Alexe B, Deselaers T, Ferrari V. "What is an object?". In: Proc. Int. Conf. Comp. Vis. Pat. Rec.; 2010.

19. Itti L, Baldi PF. Bayesian Surprise Attracts Human Attention. In: Advances in Neural Information Processing Systems; 2006.

20. Borji A, Sihite DN, Itti L. Salient Object Detection: A Benchmark. In: Proc. European Conf. Comp. Vis.; 2012.

21. Scharfenberger C, Wong A, Fergani K, Zelek JS, Clausi DA. Statistical Textural Distinctiveness for Salient Region Detection in Natural Images. In: Proc. Int. Conf. Comp. Vis. Pat. Rec.; 2013.

22. Busswell GT. How people look at pictures: A study of the psychology of perception in art. University of Chicago Press; 1935.

23. Parkhurst D, Law K, Niebur E. Modeling the role of salience in the allocation of overt visual attention. Vision Research. 2002; 42(1):107–123. doi: 10.1016/S0042-6989(01)00250-4 PMID: 11804636

24. NIST/SEMATECH. Engineering Statistics Handbook; 2012.

25. Vogel RM, Kroll CN. Low-Flow Frequency Analysis Using Probability-Plot Correlation Coefficients. Journal of Water Resources Planning and Management. 1989; 115(3):338–357. doi: 10.1061/(ASCE)0733-9496(1989)115:3(338)

26. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965; 52:591–611. doi: 10.2307/2333709

27. Schauerte B, Stiefelhagen R. How the Distribution of Salient Objects in Images Influences Salient Object Detection. In: Proc. Int. Conf. Image Process.; 2013.

28. Schauerte B, Richarz J, Plötz T, Thurau C, Fink GA. Multi-modal and multi-camera attention in smart environments. In: Proc. Int. Conf. Multimodal Interfaces; 2009.

29. Itti L, Koch C, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Trans Pattern Anal Mach Intell. 1998; 20(11):1254–1259. doi: 10.1109/34.730558

30. Harel J, Koch C, Perona P. Graph-based visual saliency. In: Advances in Neural Information Processing Systems; 2007.

31. Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. In: Proc. Int. Conf. Comp. Vis. Pat. Rec.; 2010.

32. Goferman S, Zelnik-Manor L, Tal A. Context-Aware Saliency Detection. IEEE Trans Pattern Anal Mach Intell. 2012;.

33. Hou X, Zhang L. Saliency Detection: A Spectral Residual Approach. In: Proc. Int. Conf. Comp. Vis. Pat. Rec.; 2007.

34. Hou X, Harel J, Koch C. Image Signature: Highlighting Sparse Salient Regions. IEEE Trans Pattern Anal Mach Intell. 2012; 34(1):194–201. doi: 10.1109/TPAMI.2011.146

35. Felzenszwalb PF, Huttenlocher DP. Efficient Graph-Based Image Segmentation. Int J Comput Vision. 2004; 59:167–181. doi: 10.1023/B:VISI.0000022288.19776.77

36. Schauerte B. Spectral Visual Saliency Toolbox (SViST); 2011. Available from: http://bit.ly/RAPmMk.

37. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proc. Int. Conf. Machine Learning; 2006.

38. Schauerte B, Stiefelhagen R. Look at this! Learning to Guide Visual Saliency in Human-Robot Interaction. In: Proc. Int. Conf. Intell. Robots Syst.; 2014.