

Modeling the Importance of Faces in Natural Images

Jin B.^a, Yildirim G.^a, Lau C.^a, Shaji A.^a, Ortiz Segovia M.^b and Süsstrunk S.^a

^aEPFL, Lausanne, Switzerland;

^bOcé, Paris, France

ABSTRACT

In this work we study the varying importance of faces in images. Face importance is found to be affected by the size and number of faces present. We collected a dataset of 152 face images with faces in different size and number of faces. We conducted a crowdsourcing experiment where we asked people to label the important regions of the images. Analyzing the results from the experiment, we propose a simple face-importance model, which is a 2D Gaussian function, to quantitatively represent the influence of the size and number of faces on the perceived importance of faces. The face-importance model is then tested for the application of salient-object detection. For this application, we create a new salient-objects dataset, consisting of both face images and non-face images, and also through crowdsourcing we collect the ground truth. We demonstrate that our face-importance model helps us to better locate the important, thus salient, objects in the images and outperforms state-of-the-art salient-object detection algorithms.

Keywords: Face importance, face size, number of faces, 2D Gaussian, salient-object detection

1. INTRODUCTION

Faces are important in human perception. Faces convey a wealth of information that we use to guide our social interaction.¹ From birth, humans develop their capability to detect and understand faces.² Infants as young as two days old are capable of mimicking the facial expressions of an adult.³ Adults show a preference for faces, resulting in faster processing speed for face stimuli than to non-face stimuli.⁴ The human brain even has its own special area for processing faces, called the fusiform face area.⁵

As faces are clearly important in our everyday life, they are also important in images. Several experiments demonstrate that faces are both visually salient, thus they significantly attract people’s attention,⁶ and are considered important objects in the images.⁷ Furthermore, Wang et al⁷ demonstrate that visual salience is highly correlated with the semantic importance people perceive. In this paper, we assume that importance and salience are interchangeable, where both represents the level of the viewer’s attention to any given objects. For many image processing and computer vision applications, locating the most salient regions in an image is a prerequisite. To locate the faces we can thus directly apply a face detector on the images. And, this leads to the next question: Are faces *equally* important? We show some face images in Fig. 1. Is the face in Fig. 1a equally important as the faces in Fig. 1c? Do the two faces in Fig. 1b have the same importance? We hypothesize that they are not and that the relevance of faces varies in different images. Understanding such varying importance of faces is significant for many applications such as image re-targeting⁸ and image compression.⁹

In this work, through a crowdsourcing experiment we quantitatively study the varying importance of faces in images. A simple face-importance model is proposed to quantitatively represent the influence of the size and number of faces on the perceived importance of faces. This model is further adopted for the application of salient-object detection. We first create a new salient-objects dataset where continuous importance values are obtained through crowdsourcing. To detect the salient objects in the images, our face-importance model is combined with low-level contrast features. By using our face-importance model, we show significant improvement over state-of-the-art algorithms.

We summarize our contributions as follows:

Further author information: (Send correspondence to Bin Jin)
Bin Jin: E-mail: bin.jin@epfl.ch, Telephone: +41 21 69 37604

- We introduce a face-importance model obtained by using crowdsourced ground truth. Depending on the size and number of faces in an image, our model outputs continuous importance values for faces.
- We present an algorithm that combines our face-importance model with low-level contrast-based saliency algorithms.
- We provide a new dataset, which contains both face images and non-face images, with multi-level (non-binary) ground-truth maps in pixel precision.

The paper is organized as follows: in Section 2 we discuss the related work. In Section 3 we describe the details of our crowdsourcing experiment and the face-importance model. In Section 4 we present our new salient-objects dataset, our method of salient-object detection that use the face-importance model, and the results. We conclude the paper and discuss future work in Section 5.



Figure 1. Example face images. (a) The single face is large enough to be recognized by an observer as being important. Even though there are two faces in (b), these two faces are not equally important: the larger face is much more noticeable than the smaller one. The numerous faces in (c) result in individual faces not being considered relevant.

2. RELATED WORK

The varying importance of faces in images has not been thoroughly studied. Little et al.¹⁰ address the issue of facial attractiveness based on symmetry, averageness, perceived personality and individual difference. Said et al.¹¹ discuss the role of emotions in face perception. None of them, however, focuses on quantifying the varying importance of faces in images.

Several salient-object detection algorithms recognize the importance of faces in images. Cerf et al.¹² linearly combine face-detection results with Itti's model.¹³ A face-detection prior is included in the sub-modular model proposed by Jiang et al.¹⁴ Judd et al.¹⁵ employ the support vector machine (SVM) to combine low-level features with several object detectors, including a face detector. All these algorithms directly use face-detection outputs and do not model face importance. They treat all of the detected faces with equal importance, which is not an accurate assumption, as shown in Fig. 1.

Rudoy et al.¹⁶ classify faces in images into three categories with heuristic thresholds. They reject all small faces, separate all large faces into three parts (e.g. mouth, nose and eyes) and treat all medium size faces with equal importance. Marat et al.¹⁷ show that the human gaze for faces is influenced by the size and number of faces. Their algorithm, however, does not account for the effect of face size on the importance of faces and, using a heuristic function, it lowers the saliency value of faces when multiple faces are present in a video frame. Moreover, to model the varying importance of faces, they both use heuristic functions with no experimental data, whereas we provide a function based on a crowdsourcing experiment.

3. MODELING FACE IMPORTANCE

In this work, we quantitatively study the varying importance of faces in images. During a preliminary experiment, we found that the size and number of faces are two key factors that contribute to the relevance of faces. In order to quantitatively study the relation between these two factors and the perceived face importance, we conducted

a crowdsourcing experiment on a dataset of 152 face images collected from Flickr*. These images contain either single or multiple human faces in different sizes, see Fig. 1 and Fig. 2.

The distribution of the size and number of faces in our dataset is shown in Fig. 3a. Note that the face size is the relative size of the face compared to the image size. We apply a logarithmic function on both the size and number of faces for better illustration. Therefore, in Fig. 3a, 0 on the x-axis indicates a very large face size (the same as the image size) and -4 a very small face size. Similarly, 0 on the y-axis indicates one face is present in the image, and 2 is the representation of one hundred faces. Note that there are no faces above the blue line in Fig. 3a, because the number of large faces that can exist in a single image is limited. From Fig. 3a we can see that our dataset covers a large range of the size and number of faces.

In order to obtain the importance value of each individual face, we conducted a crowdsourcing experiment on these face images. We published tasks on a crowdsourcing website† where we asked people to draw rectangles around the important regions of each image. Each time, an observer was shown only one image and was asked to label the important regions of the image. The observers were free to draw several rectangles if they thought there are several important regions in the image. Approximately 30 people labeled each image. Sample labels of one image are shown in Fig. 2a and all the labels for some example images are shown in Fig. 2b.

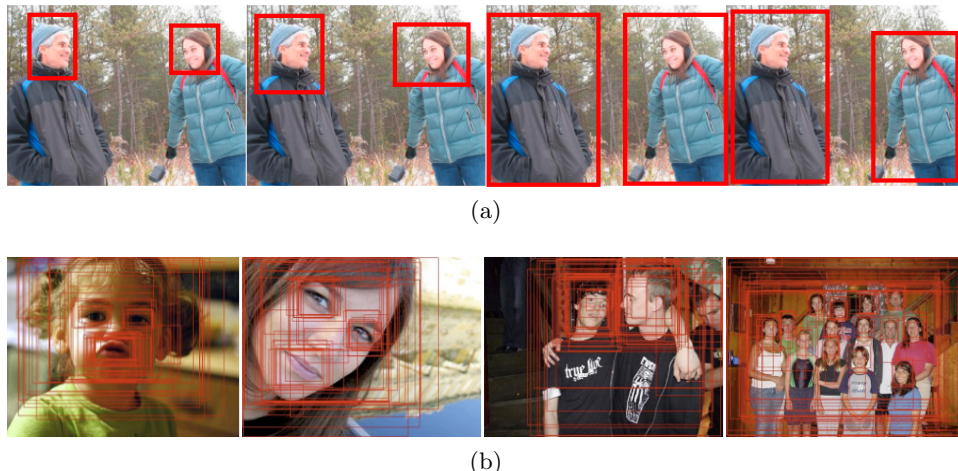


Figure 2. (a) Sample labels obtained from crowdsourcing. Observers labeled only the faces as the important region in the two images on the left. In the two images on the right, the whole body was labeled as the important region. (b) All the labels on some example images. In the two images on the left, observers labeled the whole face, as well as the eyes or mouth regions. In the third image, most observers labeled the faces as the important region of the image. In the image on the right, the whole group of people as a single object occupied most of the labels.

Based on the observers' labels, the importance value of a face is determined as

$$\psi_i^f = \frac{n_i^f}{N_i} \quad (1)$$

Here ψ_i^f is the importance value of f^{th} face in i^{th} image. N_i is the number of people who performed the labeling task for the i^{th} image. n_i^f is the number of people who specifically labeled f^{th} face in i^{th} image as an important region. Note that in two of the labels in Fig. 2a, only the faces were marked as important regions, whereas in the other two images, observers attributed importance also to the whole body. When counting n_i^f , we took this into consideration and counted only the rectangles that specifically labeled the faces.

In Fig. 3b and Fig. 3c, we plot the distribution of all ψ_i^f values over the size and number of faces, respectively. In Fig. 3d, we illustrate the joint distribution of all ψ_i^f values over the size and number of faces. As it can be seen from Fig. 3b and Fig. 3c, we can fit a two-dimensional Gaussian function to the joint distribution in Fig. 3d.

*www.flickr.com

†www.shorttask.com

This function, after normalization, is our “face-importance model” as illustrated in Fig. 3e. We show the face-importance model as a heat map in Fig. 3f. From Fig. 3f, we see that two or three faces with the size equal to 0.06 of the image size have the highest importance value. The face-importance value decreases as the number of faces increases or the size of the face decreases. Note that a single large face has relatively low importance value according to our face-importance model, because people tend to focus on the eyes or mouth rather than the whole face in this case.

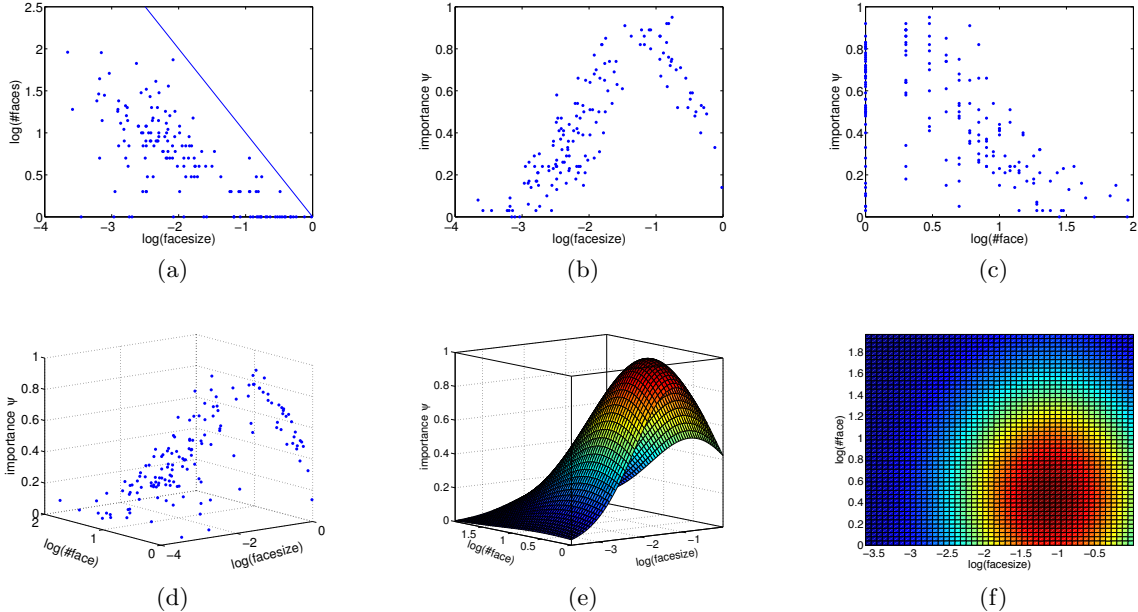


Figure 3. Results of our crowdsourcing experiment on the dataset. (a) The distribution of the size and number of faces. (b) The distribution of ψ_i^f over the face size. (c) The distribution of ψ_i^f over the number of faces. (d) The joint distribution of ψ_i^f over the size and number of faces. (e) Our face-importance model, which is a 2D Gaussian function, fitted on the distribution in (d) and then normalized between 0 and 1. (f) The face-importance model shown as a heat map.

We use the following function, denoted as G in Eqn. 2, as our face-importance model.

$$\psi_i^f \approx G(s_i^f, n_i) = e^{-\frac{(\log(s_i^f) - \mu_1)^2}{2\sigma_1^2} - \frac{(\log(n_i) - \mu_2)^2}{2\sigma_2^2}} \quad (2)$$

Here, s_i^f and n_i are the size of the f^{th} face in the i^{th} image and the number of faces in the i^{th} image, respectively. Note that s_i^f is the relative size compared to the size of the image, therefore it is between 0 and 1. The parameters of the Gaussian fit are $\mu_1 = -1.172$, $\mu_2 = 0.4308$, $\sigma_1 = 0.9701$, $\sigma_2 = 0.7799$, and the base of the logarithm is 10. The R^2 value of our 2D Gaussian fit is equal to 0.8567, which demonstrates that our model accurately describes the experimental importance of faces, namely ψ_i^f .

We use a 2D Gaussian function as it quantitatively explains the influence of the size and number of faces on the perceived face importance. Moreover, the 2D Gaussian function is simple and thus could be easily incorporated into other algorithms.

4. SALIENT-OBJECT DETECTION

To test the effectiveness of our face-importance model, we propose a simple yet effective algorithm for salient object detection, by combining our face-importance model with the low-level contrast feature.

4.1 Dataset

There are several available datasets for evaluating salient object detection algorithms such as the ASD dataset,¹⁸ the MSRA dataset¹⁹ and the PASCAL dataset.²⁰ None of these datasets, however, are suitable for evaluating our method, because we want to test the effectiveness of our face-importance model and these datasets do not include a sufficient amount of face images. Therefore, we build a new dataset consisting of 800 images, 632 of which include human faces. The largest dimension of each image is between 1000 to 1024 pixels. The faces in this dataset cover a large variance of the size and number of faces. Note that the 152 images that we used previously to model the face importance are not included in this dataset. Some sample images from our evaluation dataset are shown in Fig. 4.

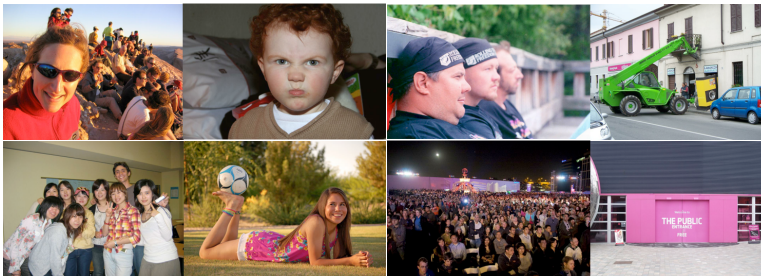


Figure 4. Sample images from our evaluation dataset. In addition to images with human faces, our evaluation dataset also includes images of salient objects.

We also perform a crowdsourcing experiment on this dataset. An illustration of the rectangle labeling is given in the first row of Fig. 5. In the second row, to reveal the most salient regions, we accumulate the rectangles and create a weight map. The intensity value of the weight map represents the importance of the region. We further provide manually labeled pixel-level ground truth for the important objects. For this, we first binarize the weight map by using Ostu’s method,²¹ producing the foreground and background areas in the third row in Fig. 5. Then we manually outline all the salient objects in the foreground area. Note that we label the face and the human body as separate objects, as long as there are rectangles specifically drawn on the face. Otherwise the human body and the face are labeled as a whole. The same principle is applied to the eyes, the mouth and the face. The final ground truth, in the last row, is created by assigning the mean value of the salient-object regions in the weight map to the corresponding region in the ground truth. For example, in the second column of Fig. 5, the saliency value for the eyes in the ground truth (the last row) is the average intensity value of the corresponding regions in the weight map (second row). Instead of providing a binary ground truth,^{18,22} we provide a multi-level ground truth, which not only represents the clear outline of the salient objects, but also emphasizes the diversity of importance among salient objects.

4.2 Our Method

Using our face-importance model, we propose a novel feature, the “face-importance map”, to represent the varying importance of faces. High value in the “face-importance map” implies high importance of the face. This “face-importance map” is combined with a low-level contrast-based feature for robustly locating the salient objects. The flowchart of this algorithm is shown in Fig. 6.

4.2.1 Face-Importance Map

In order to generate a precise face-importance map, we first over segment an image into 1000 superpixels by using the SLIC algorithm.²³ This step assures the internal consistency and external separation of image regions. We then use Zhu and Ramanan’s²⁴ face detector, that robustly detects faces larger than 80×80 pixels. For smaller faces, we employ Viola-Jones.²⁵ We outline the face region by considering superpixel information and the face detection output, Fig. 7. In order to form the final face-importance map, we use the following equation:

$$\mathcal{F}_i = \sum_{f=1}^{n_i} G(s_i^f, n_i) \cdot Q^f \quad (3)$$

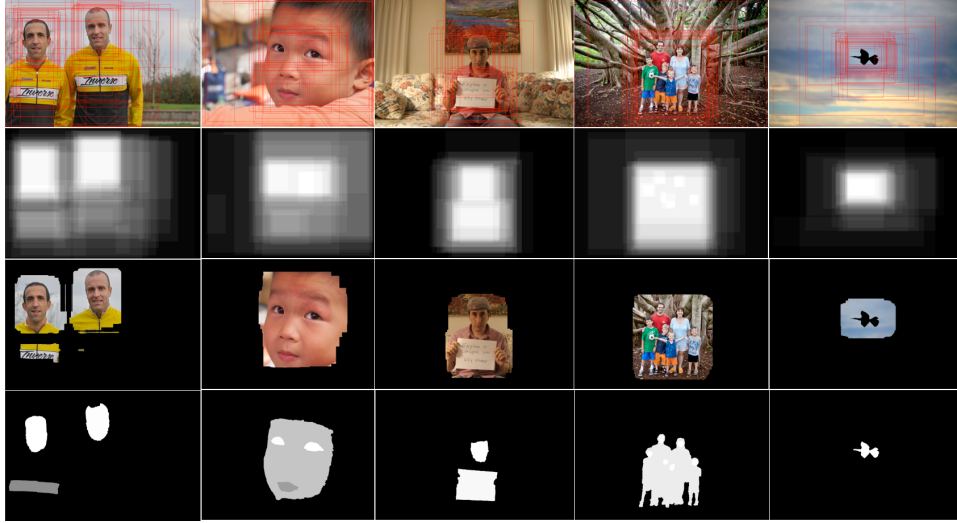


Figure 5. The first row: the combination of all the labels from crowd-sourcing; the second row: the corresponding weight map; the third row: the foreground area using Ostu's threshold; the last row: our multi-level ground truth

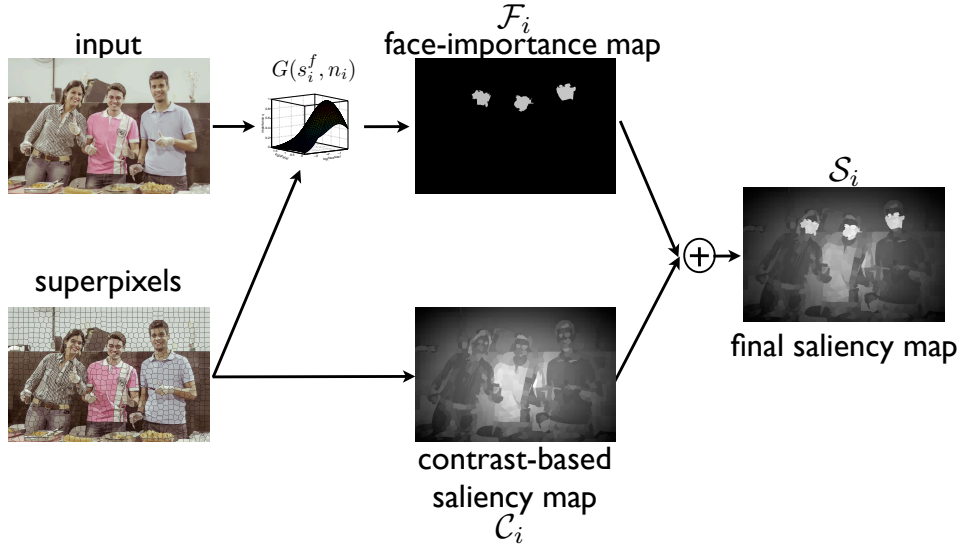


Figure 6. The flowchart of our algorithm. We first over segment an input image into superpixels. A contrast-based saliency map is computed using superpixels. The face-importance map is calculated using both the input image and its superpixels through our face-importance model. The final saliency map is the addition of the two maps.

Here, \mathcal{F}_i is the face-importance map of i^{th} image, Q^f is a binary image with the same size as i^{th} image, where only the superpixels for the f^{th} face are set to 1.

4.2.2 Final Saliency Map

The final saliency map is the combination of the high-level face-importance map with low-level contrast:

$$S_i = \mathcal{F}_i + C_i \quad (4)$$

Here, S_i is the final saliency map of i^{th} image and C_i is the low-level contrast-based saliency map of i^{th} image, which is explained in the following paragraphs.

We chose to model the combination of low-level contrast and face importance with an addition because it is biologically plausible. As with low-level contrast, faces are also instantly perceived and neurologically significant.

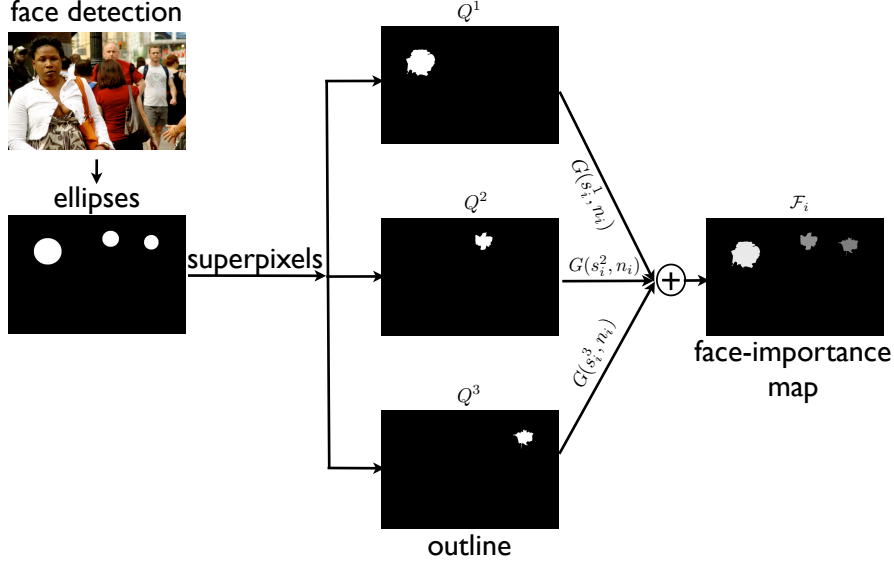


Figure 7. Building a face-importance map. We first fit ellipses into the face-detection rectangles and recover the superpixels that have more than half of their area inside the ellipses. These superpixel groups are treated as face areas, and then weighted by our face-importance model. Note that the gray-level in the face-importance map represents the importance.

We know that in the human brain there is a specific area for processing faces, called the fusiform face area.⁵ Therefore, we model the union of the two neural pathways, one for low-level contrast and the other for face importance, with addition.

4.2.3 Contrast-Based Saliency Map

Color-contrast features have been effectively used for salient object detection in previous works:^{13,26,27} they compute either local center-surround contrast¹³ or global contrast²⁶ or a combination of both.²⁷ In order to compute the low-level color contrast, we follow an approach that is similar to the one used by Perazzi et al.²⁷

Our method differs from their method²⁷ because we use more than one single parameter for the spatial support of the contrast measure. We sweep this parameter to calculate the contrast both locally and globally. We also take into account the size of the superpixel (the difference from a large superpixel results in high contrast). In addition, to obtain more accurate saliency maps we multiply our saliency map with the spatial compactness and the adaptive center prior measures. We compute the contrast value of a superpixel as follows:

$$\zeta(S_k, \delta, \alpha) = \sum_{p=1}^N n_p \cdot |\Lambda_p^\alpha - \Lambda_k^\alpha| \cdot \exp\left(-\frac{1}{\delta} \cdot \frac{\|\mathbf{p}_p - \mathbf{p}_k\|_2}{\sqrt{m^2 + n^2}}\right) \quad (5)$$

Here, $\zeta(S_k, \delta, \alpha)$ is the contrast value of superpixel S_k using δ that adjusts the spatial support of the contrast measure, and α that represents one of the channels in CIELab color space. $\{\Lambda_k^\alpha | k = 1, 2, \dots, N_S, \alpha = L, a, b\}$ are the average color values of the superpixels in the CIELab color space, N is the number of superpixels, n_p is the size of superpixel S_p , specifically the number of pixels in the superpixel, \mathbf{p}_k is the spatial position vector of superpixel S_k , m and n are the width and height of the input image, respectively.

For a certain δ and α , we compute the contrast for all superpixels, thus forming a saliency map $\eta(x, y, \delta, \alpha)$, where x and y represent the spatial coordinates. The final contrast-based saliency map is computed as follows:

$$\mathcal{C}_i = \sum_{\delta \in \Delta} \sum_{\alpha=L,a,b} K(\delta, \alpha) \cdot \mathbf{P}(x, y, \delta, \alpha) \cdot \eta(x, y, \delta, \alpha) \quad (6)$$

Here, Δ is a set of values to measure both local and global contrast. We experimentally set $\Delta = \{0.01, 0.05, 0.09, 0.1, 0.5, 0.9, 1, 4\}$ in our implementation. K is the spatial compactness measure and \mathbf{P} is the adaptive center prior

measure. In order to compute K and \mathbf{P} , we first compute the following parameters:

$$\mu_x(\delta, \alpha) = \frac{1}{m \cdot n} \sum_x \sum_y x \cdot \bar{\eta}(x, y, \delta, \alpha) \quad (7)$$

$$\tilde{\mu}_x(\delta, \alpha) = \frac{1}{m \cdot n} \sum_x \sum_y x \cdot (1 - \bar{\eta}(x, y, \delta, \alpha)) \quad (8)$$

$$\sigma_x^2(\delta, \alpha) = \frac{1}{m \cdot n} \sum_x \sum_y (x - \mu_x)^2 \cdot \bar{\eta}(x, y, \delta, \alpha) \quad (9)$$

$$\tilde{\sigma}_x^2(\delta, \alpha) = \frac{1}{m \cdot n} \sum_x \sum_y (x - \tilde{\mu}_x)^2 \cdot (1 - \bar{\eta}(x, y, \delta, \alpha)) \quad (10)$$

Here, $\bar{\eta}$ is obtained by mapping η between 0 and 1. Similar equations are used for the computation of μ_y , σ_y^2 , $\tilde{\mu}_y$, and $\tilde{\sigma}_y^2$. Here, μ_x and μ_y measure the center of the mass of the saliency map, $\eta(x, y, \delta, \alpha)$, and σ_x^2 and σ_y^2 are related to the level of distribution (or compactness) of the saliency map. $\tilde{\sigma}_x^2$ and $\tilde{\sigma}_y^2$ are used to measure the distribution of non-salient regions. Then the compactness K and the adaptive prior \mathbf{P} is computed as (δ and α are dropped for convenience):

$$K(\delta, \alpha) = \exp\left(-u \cdot \frac{\sigma_x^2 \cdot \sigma_y^2}{\tilde{\sigma}_x^2 \cdot \tilde{\sigma}_y^2}\right), u = 4 \quad (11)$$

$$\mathbf{P}(x, y, \delta, \alpha) = \exp\left(-\frac{(x - \mu_x)^2}{v \cdot \sigma_x^2} - \frac{(y - \mu_y)^2}{v \cdot \sigma_y^2}\right), v = 12 \quad (12)$$

4.3 Results

We compare our method to eight state-of-the-art salient-object detection algorithms, including low-level contract based models (AMC,²⁸ CH,²⁹ GBMR³⁰) and models that include a face detector (SMVJ,¹² LR,³¹ Judd et al.,¹⁵ Borji et al.,³² SC³³). Note that the ground truth in our dataset has multi-level values. For machine-learning-based methods, such as Judd et al.,¹⁵ Borji et al.³² and SC,³³ we train a new model by using our dataset and adopt linear regression with 10-fold cross-validation.

The precision-recall curves in Fig. 8a show that our algorithm significantly outperforms all eight state-of-the-art methods. As our ground truth is not binary, to compute the precision and recall values, we binarize both the saliency maps and the ground truth by sweeping thresholds from 0 to 255. The high-importance regions remain part of the ground truth for a large portion of the thresholds, whereas low-importance regions disappear at a certain threshold. By using this method, we are able to evaluate the varying importance of salient objects. Also due to this factor, the precision values of all of the methods decrease as recall approaches zero, because the salient objects in the ground truth are similarly decreasing.

As expected, the methods without a face detector, e.g. AMC,²⁸ CH,²⁹ and GBMR,³⁰ do not perform well. Judd et al.¹⁵ Borji et al.,³² and SC³³ report quite similar performances because they adopt similar low-level and high-level features. Unlike our algorithm, theirs directly incorporate binary face detection output as a feature and thus fail to model the varying importance of faces. Therefore, our method performs significantly better. In Fig. 8b we illustrate the precision, the recall, and the F-measure by using the same adaptive thresholding as in Achanta et al.¹⁸ Our method achieves the highest F-measure among all the methods, thus demonstrating a balance between the precision and recall value.

Fig. 9 shows sample saliency maps from the nine saliency methods. Our saliency algorithm computes robust saliency maps for both face images and non-face images. Our method locates the important faces in the image (the second row in Fig. 9) as well as the high contrast region (the sixth row in Fig. 9).

4.4 The Effectiveness of the Face-Importance Map

To evaluate the effectiveness of our novel face-importance map, we add it to different state-of-the-art salient-object detection methods. We use linear summation with equal weights to combine the face-importance map with

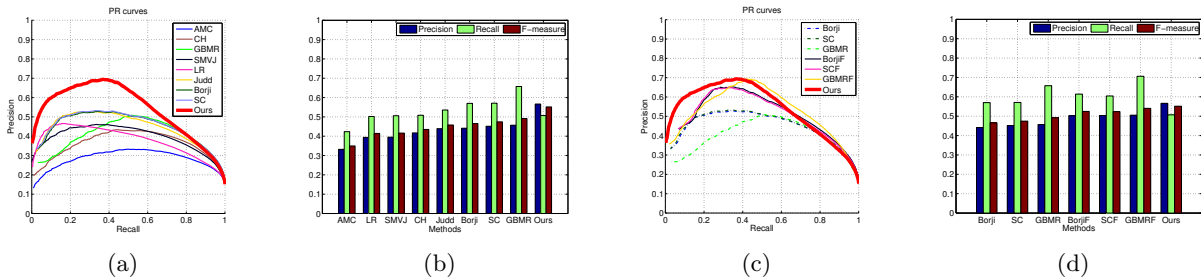


Figure 8. (a) (b) the PR curves and F-measure for all the methods using multi-level ground truth. Our method significantly outperforms all state-of-the-art methods. (c) (d) Comparison of performance for methods with face-importance map (GBMRF, BrojiF and SCF) and methods without face-importance map (GBMR, Borji and SC). Including face-importance map improves the performance

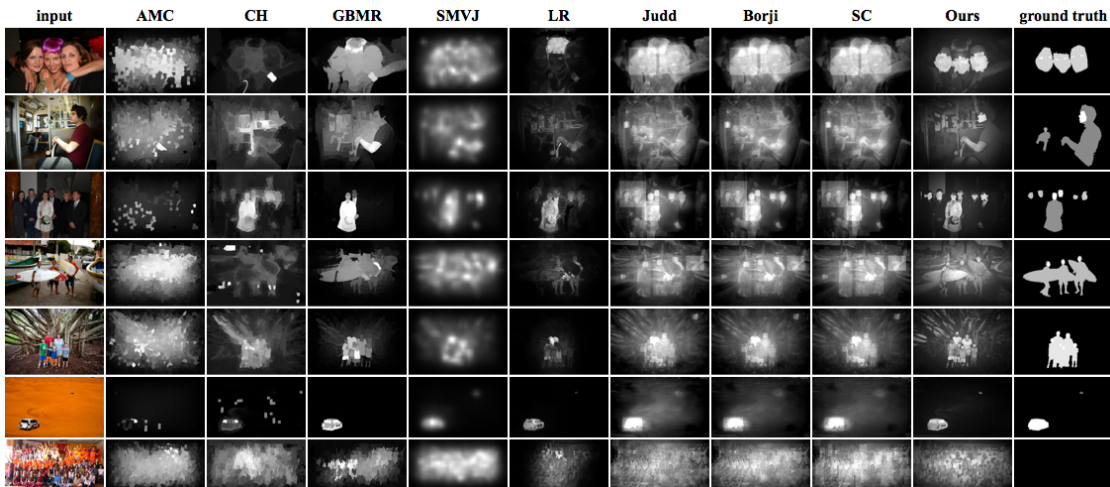


Figure 9. The saliency maps from different saliency methods. Our algorithm consistently achieves better results. The ground truth for the last row is empty, because the labeling in the crowd-sourcing experiment is highly diverse, which implies no specific salient region in this image.

a low-level contrast based method, GBMR;³⁰ hence we obtain “GBMRF”. The combined saliency map, GBMRF, significantly outperforms its original results, GBMR, as shown in Fig. 8c. We also replace the face-detection feature, used in SC³³ and Borji et al.,³² with our face-importance map, and we train the model again using regression and a 10-fold cross-validation. As illustrated in Fig. 8c and 8d, the new saliency maps, represented as SCF and BorjiF, report remarkable improvement compared to their original algorithm. Due to the accuracy of our contrast-based saliency maps, our algorithm still outperforms the modified GBMRF, SCF, and BorjiF methods, yet by a much smaller degree than before.

5. CONCLUSION AND FUTURE WORK

In this work we have quantitatively studied the varying importance of faces in images through crowdsourcing. We propose a face-importance model, a 2D Gaussian function, to represent the effect of the size and number of faces on the perceived importance of faces. The face-importance model is combined with low-level contrast to form a salient-object detection method. Evaluated on our salient-objects dataset, which we provide with multi-level pixel-wise ground truth, our method notably outperforms state-of-the-art salient object detection algorithms.

We have investigated two major factors that affect the perceived importance of faces in images. The influence of other factors, like pose, occlusion or gender, need to be further studied and combined with our face-importance model. Moreover, the proposed model enables us to incorporate the varying face importance into many more applications, such as image cropping⁸ and image collage.³⁴

REFERENCES

- [1] Jeffery, L. and Rhodes, G., “Insights into the development of face recognition mechanisms revealed by face aftereffects,” *British Journal of Psychology* **102**(4), 799–815 (2011).
- [2] Little, A. C., Jones, B. C., and DeBruine, L. M., “The many faces of research on face perception,” *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**(1571), 1634–1637 (2011).
- [3] Field, T. M., Woodson, R., Greenberg, R., and Cohen, D., “Discrimination and imitation of facial expressions by neonates,” *Annual Progress in Child Psychiatry & Child Development* **16**, 119–125 (1983).
- [4] Dawson, G., Webb, S. J., and McPartland, J., “Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies,” *Developmental neuropsychology* **27**(3), 403–424 (2005).
- [5] Kanwisher, N., McDermott, J., and Chun, M. M., “The fusiform face area: a module in human extrastriate cortex specialized for face perception,” *The Journal of Neuroscience* **17**(11), 4302–4311 (1997).
- [6] Zhao, Q. and Koch, C., “Learning a saliency map using fixated locations in natural scenes,” *Journal of vision* **11**(3), 9 (2011).
- [7] Wang, J., Chandler, D. M., and Le Callet, P., “Quantifying the relationship between visual salience and visual importance,” *IS&T/SPIE Electronic Imaging*, 75270K–75270K, International Society for Optics and Photonics (2010).
- [8] Avidan, S. and Shamir, A., “Seam carving for content-aware image resizing,” *ACM Transactions on Graphics* **26**(3) (2007).
- [9] Stella, X. Y. and Lisin, D. A., “Image compression based on visual saliency at individual scales,” 157–166, Springer (2009).
- [10] Little, A. C., Jones, B. C., and DeBruine, L. M., “Facial attractiveness: evolutionary based research,” *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**(1571), 1638–1659 (2011).
- [11] Said, C. P., Haxby, J. V., and Todorov, A., “Brain systems for assessing the affective value of faces,” *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**(1571), 1660–1670 (2011).
- [12] Cerf, M., Harel, J., Einhäuser, W., and Koch, C., “Predicting human gaze using low-level saliency combined with face detection,” *NIPS* (2007).
- [13] Itti, L., Koch, C., and Niebur, E., “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(11), 1254–1259 (1998).
- [14] Jiang, Z. and Davis, L. S., “Submodular salient region detection,” *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2043–2050 (2013).
- [15] Judd, T., Ehinger, K., Durand, F., and Torralba, A., “Learning to predict where humans look,” in [*Proceedings of IEEE International Conference on Computer Vision*], 2106–2113 (2009).
- [16] Rudoy, D., Goldman, D. B., Shechtman, E., and Zelnik-Manor, L., “Learning video saliency from human gaze using candidate selection,” *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 1147–1154, IEEE (2013).
- [17] Marat, S., Guyader, N., and Pellerin, D., “Gaze prediction improvement by adding a face feature to a saliency model,” *Recent advances in signal processing*, 195–210 (2009).
- [18] Achanta, R., Hemami, S., Estrada, F., and Süsstrunk, S., “Frequency-tuned salient region detection,” *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1597–1604 (2009).
- [19] Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., and Li, S., “Salient object detection: A discriminative regional feature integration approach,” *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2083–2090 (2013).
- [20] Zou, W., Kpalma, K., Liu, Z., Ronsin, J., et al., “Segmentation driven low-rank matrix recovery for saliency detection,” *Proceedings of British Machine Vision Conference*, 1–13 (2013).
- [21] Otsu, N., “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man and Cybernetics* **9**(1), 62–66 (1979).
- [22] Movahedi, V. and Elder, J. H., “Design and perceptual validation of performance measures for salient object segmentation,” *Proceedings of IEEE Computer Vision and Pattern Recognition Workshops*, 49–56 (2010).

- [23] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S., "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(11), 2274–2282 (2012).
- [24] Zhu, X. and Ramanan, D., "Face detection, pose estimation, and landmark localization in the wild," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* , 2879–2886 (2012).
- [25] Viola, P. and Jones, M., "Rapid object detection using a boosted cascade of simple features," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **1**, 511–518 (2001).
- [26] Cheng, M.-M., Zhang, G.-X., Mitra, N. J., Huang, X., and Hu, S.-M., "Global contrast based salient region detection," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* , 409–416, IEEE (2011).
- [27] Perazzi, F., Krahenbuhl, P., Pritch, Y., and Hornung, A., "Saliency filters: Contrast based filtering for salient region detection," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* , 733–740 (2012).
- [28] Jiang, B., Zhang, L., Lu, H., Yang, M.-H., and Yang, C., "Saliency detection via absorbing markov chain," *Proceedings of IEEE International Conference on Computer Vision* (2013).
- [29] Li, X., Li, Y., Shen, C., Dick, A., and van den Hengel, A., "Contextual hypergraph modelling for salient object detection," *Proceedings of IEEE International Conference on Computer Vision* (2013).
- [30] Yang, C., Zhang, L., Lu, H., Ruan, X., and Yang, M.-H., "Saliency detection via graph-based manifold ranking," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* , 3166–3173 (2013).
- [31] Shen, X. and Wu, Y., "A unified approach to salient object detection via low rank matrix recovery," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* , 853–860 (2012).
- [32] Borji, A., "Boosting bottom-up and top-down visual features for saliency estimation," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* , 438–445 (2012).
- [33] Karthikeyan, S., Jagadeesh, V., and Manjunath, B. S., "Learning top down scene context for visual attention modelling in natural images," *Proceedings of IEEE International Conference on Image Processing* (2013).
- [34] Goferman, S., Tal, A., and Zelnik-Manor, L., "Puzzle-like collage," *Computer Graphics Forum* **29**(2), 459–468, Wiley Online Library (2010).