Hierarchical Image Saliency Detection on Extended CSSD

Jianping Shi, *Student Member, IEEE* Qiong Yan, *Member, IEEE* Li Xu, *Member, IEEE*
Jiaya Jia, *Senior Member, IEEE*

Abstract—Complex structures commonly exist in natural images. When an image contains small-scale high-contrast patterns either in the background or foreground, saliency detection could be adversely affected, resulting erroneous and non-uniform saliency assignment. The issue forms a fundamental challenge for prior methods. We tackle it from a scale point of view and propose a multi-layer approach to analyze saliency cues. Different from varying patch sizes or downsizing images, we measure region-based scales. The final saliency values are inferred optimally combining all the saliency cues in different scales using hierarchical inference. Through our inference model, single-scale information is selected to obtain a saliency map. Our method improves detection quality on many images that cannot be handled well traditionally. We also construct an extended Complex Scene Saliency Dataset (ECSSD) to include complex but general natural images.

Index Terms—saliency detection, region scale

✦

1 INTRODUCTION

Regarding saliency, each existing method mainly focuses on one of the following tasks – i.e., eye-fixation prediction, image-based salient object detection and objectness estimation. Among them, image-based salient object detection [9], [6], [5], [47], [44] is an important stream, which can benefit several applications including detection [19], classification [37], retrieval [17], and object co-segmentation [7], for optimizing and saving computation. The goal is to detect and segment out important regions from natural images.

By defining pixel/region uniqueness in either local or global context, existing image salient object detection methods can be classified to two categories. Local methods [1], [24] rely on pixel/region difference in the vicinity, while global methods [9], [34], [44] rely mainly on color uniqueness statistically.

Albeit many methods were proposed, a few common issues still endure. They are related to complexity of patterns in natural images. A few examples are shown in Fig. 1. For the first two examples, the boards containing characters are salient foreground objects. But the results in (b), produced by a previous local method, only highlight a few edges that scatter in the image. The global method results in (c) also cannot clearly distinguish among regions. Similar challenge arises when the background is with complex patterns, as shown in the last example of Fig. 1. The yellow flowers lying on grass stand out by previous methods.

● *Jianping Shi and Jiaya Jia are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong. The contact is in http://www.cse.cuhk.edu.hk/leojia/.*

But they are actually part of the background when viewing the picture as a whole.

These examples are not special, and exhibit one common problem – that is, *when objects contain salient small-scale patterns, saliency could generally be misled by their complexity*. Given texture existing in many natural images, this problem cannot be escaped. It easily turns extracting salient objects to finding cluttered fragments of local details, complicating detection and making results not usable in, for example, object recognition [43], where connected regions with reasonable sizes are favored.

Aiming to solve this notorious and universal problem, we propose a hierarchical framework, to analyze saliency cues from multiple levels of structure, and then integrate them for the final saliency map through hierarchical inference. Our framework finds foundation from studies in psychology [31], [27], which show the selection process in human attention system operates from more than one levels, and the interaction between levels is more complex than a feed-forward scheme. Our multi-level analysis helps deal with salient small-scale structures. The hierarchical inference plays an important role in fusing information to get accurate saliency maps.

Our contributions in this paper also include 1) a new measure of region scales, which is compatible with human perception on object scales, and 2) extension of Complex Scene Saliency Dataset (CSSD), which contains 1000 challenging natural images for saliency detection. Our method yields improvement over others on the new extended CSSD dataset as well as other benchmark datasets.

This manuscript extends the conference version [46] with the following major differences. First, we provide

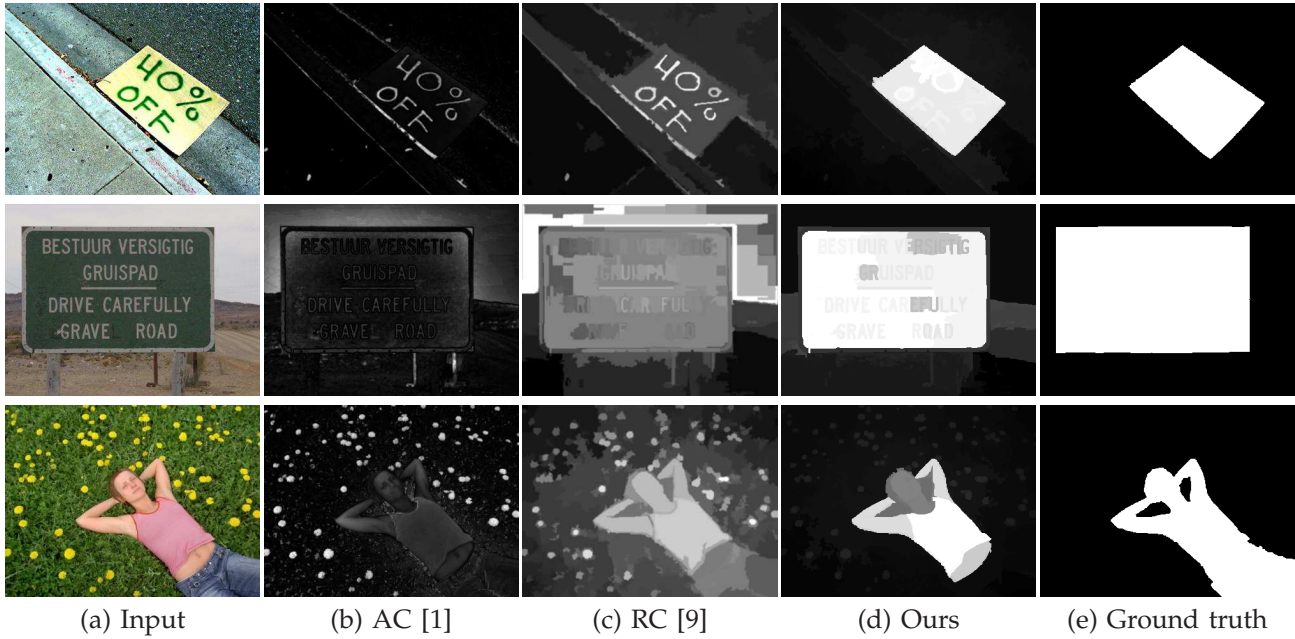| (a) Input | (b) AC [1] | (c) RC [9] | (d) Ours | (e) Ground truth |

Fig. 1. Saliency detection with structure confusion. Small-scale strong details easily influence the process and cause erroneous results.

more analysis on region scale computation and region merge. Second, we build a new hierarchical inference model with a local consistency scheme, which leads to more natural saliency results compared to previous tree-structured model. Further, we build an extended Complex Scene Saliency Dataset (ECSSD) with more challenging natural images.

The rest of the paper is organized as follows. Section 2 reviews literature in saliency detection. In Section 3, we introduce our hierarchical solutions for saliency detection. We conduct experiments in Section 4 and conclude this paper in Section 5.

2 RELATED WORK

Saliency analysis generally follows eye fixation location and object-based attention formation [33]. Eye fixation location methods physically obtain human attention shift continuously with eye tracking, while object-based approaches aim to find salient objects from the input. The salient object detection is further extended to "objectness estimation" in object recognition. Both of them are important and benefit different applications in high-level scene analysis. Extensive review was provided in [42], [5]. Below we discuss a few. Note that in this paper we only address image-based salient object detection problem.

Eye Fixation Prediction and Objectness Eye fixation methods compute a saliency map to match eye movement. The early method [20] used an image pyramid to calculate pixel contrast based on color and orientation features. Ma and Zhang [30] directly computed center-surrounding color difference in a fixed neighborhood for each pixel. Harel *et al.* [16] proposed a method to non-linearly combine local uniqueness maps from different feature channels to concentrate conspicuity. Judd *et al.* [22] combined high level human detector and center priors into eye fixation prediction. Borji and Itti [4] considered local and global image patch rarities in two color space, and fuse information.

Objectness is anther direction on saliency detection. It is to find potential objects [10] based on the low level clues of the input image independent of their classes. In particular, it measures whether a particular bounding box represents an object. Endres and Hoiem [12] proposed an object proposal method based on segmentation. Alexe *et al.* [3] integrated saliency clues for object prediction. Cheng *et al.* [10] developed a gradient feature for objectness estimation.

Salient Object Detection Salient object detection, different from above problems, segments salient objects out. Local methods extract saliency features regarding a neighboring region. In [29], three patch-based features are learned and connected via conditional random field. Achanta *et al.* [1] defined local pixel saliency using local luminance and color. This method needs to choose an appropriate surrounding patch size. Besides, high-contrast edges are not necessarily in the foreground as illustrated in Fig. 1.

Global methods mostly consider color statistics. Zhai and Shah [48] introduced image histograms to calculate color saliency. To deal with RGB color, Achanta *et al.* [2] provided an approximate by subtracting the

(a) Input image

(b) Final saliency map

(c) Image layers

(d) Cue maps

(e) Inference

$E_H(s_i^k, s_j^{k+1})$

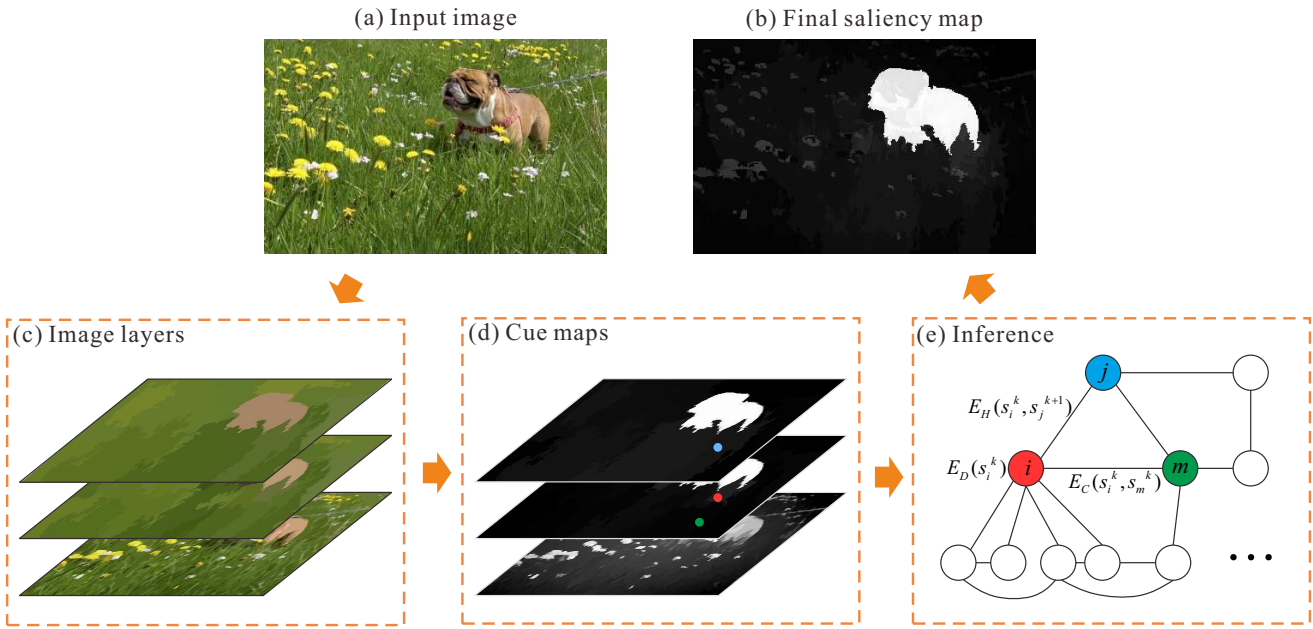$E_D(s_i^k)$

$E_C(s_i^k, s_m^k)$

Fig. 2. An overview of our hierarchical framework. We extract three image layers from the input, and then compute saliency cues from each of these layers. They are finally fed into a hierarchical inference model to get the final results.

average color from the low-pass filtered input. Cheng *et al.* [9] extended the histogram to 3D color space. These methods find pixels/regions with colors much different from the dominant one, but do not consider spatial locations. To compensate the lost spatial information, Perazzi *et al.* [34] measured the variance of spatial distribution for each color. Global methods have their difficulty in distinguishing among similar colors in both foreground and background. Recent methods exploit background smoothness [38], [44]. Note smooth structure assumption could be invalid for many natural images, as explained in Section 1.

High-level priors were also used based on common knowledge and experience. Face detector was adopted in [14], [38]. The concept of center bias – that is, image center is more likely to contain salient objects than other regions – was employed in [29], [38], [44]. In [38], it is assumed that warm colors are more attractive to human. Learning techniques are popular in several recent methods [23], [35], [39]. Unique features or patterns are learned from a large set of labeled images or a single image in an unsupervised manner. Li *et al.* [28] links the eye fixation prediction and salient object detection via segmentation candidates.

Prior work does not consider the situation that locally smooth regions could be inside a salient object and globally salient color, contrarily, could be from the background. These difficulties boil down to the same type of problems and indicate that saliency is ambiguous in one single scale. As image structures exhibit different characteristics when varying resolutions, they should be treated differently to embody

diversity. Our hierarchical framework is a unified one to address these issues.

3 HIERARCHICAL FRAMEWORK

Our method starts from layer extraction, by which we extract images of different scales from the input. Then we compute saliency cues for each layer, which are then used to infer the final saliency confidence in a local consistent hierarchical inference model. The framework is illustrated in Fig. 2.

3.1 Image Layer Extraction

Image layers, as shown in Fig. 2(c), are coarse representation of the input with different degrees of details, balancing between expression capability and structure complexity. The layer number is fixed to 3 in our experiments. In the bottom level, finest details such as flower are retained, while in the top level large-scale structures are produced.

3.1.1 Layer Generation

To produce the three layers, we first generate an initial over-segmentation as illustrated in Fig. 3(b) by the watershed-like method [15]. For each segmented region, we compute a scale value, where the process is elaborated on in the next subsection. They enable us to apply an iterative process to merge neighboring segments. Specifically, we sort all regions in the initial map according to their scales in an ascending order. If

(a) Input   (b) Over-segmentation   (c) Layer $\mathcal{L}^1$ (d) Layer $\mathcal{L}^2$ (e) Layer $\mathcal{L}^3$
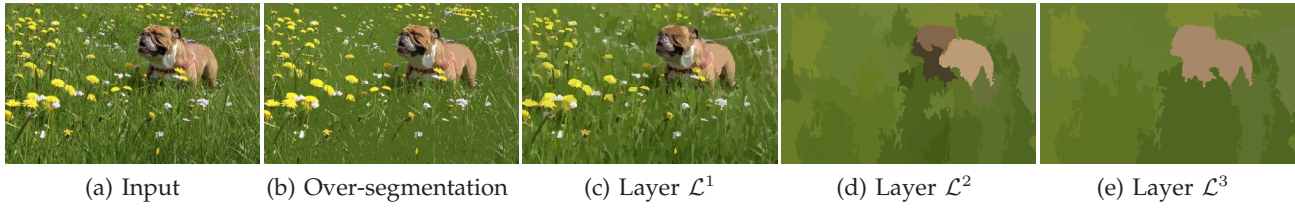
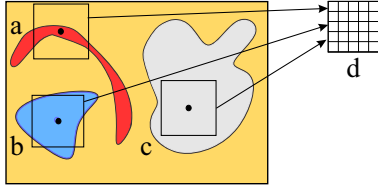Fig. 3. Region-merge results under different scales.

Fig. 4. Our region scale is defined as the largest square that a region can contain. In this illustration, the scales of regions a and b are less than 5, and that of c is larger than 5.

a region scale is below a selected threshold, we merge it to its nearest region, in terms of average CIELUV color distance, and update its scale. We also update the color of the region as their average color. After all regions are processed, we take the resulting region map as the bottom layer, denoted as $\mathcal{L}^1$. The superscript here indexes the first layer among the three ones we operate. In what follows without further explanation, the super-script is the layer index.

The middle and top layers $\mathcal{L}^2$ and $\mathcal{L}^3$ are generated similarly from $\mathcal{L}^1$ and $\mathcal{L}^2$ with larger scale thresholds. In our experiment, we set thresholds for the three layers as $\{5, 17, 33\}$ for typical 400×300 images. Three layers are shown in Fig. 3(c)-(e). More details on scale computation and region merge are described in the following subsections. Note a region in the middle or top layer embraces corresponding ones in the lower levels. We use the relationship for saliency inference described in Section 3.3.

3.1.2 Region Scale Definition

In methods of [11], [13] and many others, the region size is measured by the number of pixels. Our research and extensive experiments suggest this measure could be wildly inappropriate for processing and understanding general natural images. In fact, a large pixel number does *not* necessarily correspond to a large-scale region in human perception.

An example is shown in Fig. 4. Long curved region a contains many pixels. But it is not regarded as a large region in human perception due to its high inhomogeneity. Region b could look bigger although its pixel number is not larger. With this fact, we define a new *encompassment* scale measure based on shape
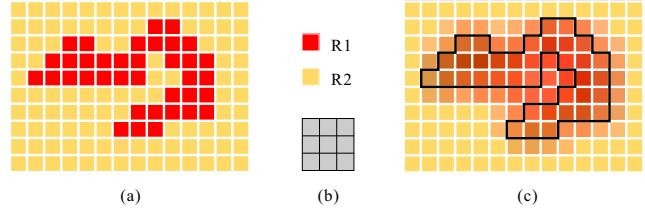

Fig. 5. Efficient computation of scale transform. (a) Initial region map. (b) Map labels and the box filter. (c) Filtered region map. As shown in (c), all colors in R1 are updated compared to the input, indicating a scale smaller than 3.

uniformities and use it to obtain region sizes in the merging process.

Definition *Region R encompassing region R' means there exists at least one location to put R' completely inside R, denoted as $R' \subseteq R$.*

With this relation, we define the scale of region R as

$$\text{scale}(R) = \arg\max_t \{R_{t \times t} | R_{t \times t} \subseteq R\}, \quad (1)$$

where $R_{t \times t}$ is a $t \times t$ square region. In Fig. 4, the scales of regions a and b are smaller than 5 while the scale of c is above it.

3.1.3 Efficient Algorithm to Compute Region Scale

To determine the scale for a region, naive computation following the definition in Eq. (1) needs exhaustive search and comparison, which could be costly. In fact, in the merging process in a level, we only need to know whether the scale of a region is below the given threshold t or not. We resort to a fast method by spatial convolution.

Given a map M with each pixel labeled by its region index in the region list $\mathcal{R}$, we apply a box filter k_t of size $t \times t$, which produces a blurred map $k_t \circ M$ ($\circ$ denotes 2D convolution).

With computation of absolute difference $D_t = |M - k_t \circ M|$, we screen out regions in $\mathcal{R}$ with their scales smaller than t. The scale for a region R_i is smaller than t if and only if

$$\left(\min_y \{D_t(y) | y \in R_i\} \right) > 0, \quad (2)$$

Algorithm 1 Scale Estimation

1: **input:** Region list $\mathcal{R}$, scale threshold t
2: Create a map M with each pixel labeled by its region index in $\mathcal{R}$;
3: Create a box filter k_t of size $t \times t$;
4: $D_t \leftarrow |M - k_t \circ M|$;
5: $\mathcal{R}_t \leftarrow \emptyset$;
6: **for** each region R_i in $\mathcal{R}$ **do**
7: $\quad x \leftarrow \min_y \{D_t(y) | y \in R_i\}$;
8: $\quad$ If $x > 0$ then $\mathcal{R}_t \leftarrow \mathcal{R}_t \bigcup \{R_i\}$;
9: **end for**
10: **output:** Region list $\mathcal{R}_t$

Algorithm 2 Region Merge

1: **input:** Region list $\mathcal{R}$, scale threshold t
2: **repeat**
3: $\quad$ Get region list $\mathcal{R}_t$ by Algorithm 1;
4: $\quad$ **for** each region R_i in $\mathcal{R}_t$ **do**
5: $\quad\quad$ Find the neighboring region $R_j \in \mathcal{R}$ with the minimum Euclidian distance to R_i in CIELUV color space;
6: $\quad\quad$ Merge R_i to R_j;
7: $\quad\quad$ Set the color of R_j to the average of R_i and R_j;
8: $\quad$ **end for**
9: **until** $\mathcal{R}_t = \emptyset$
10: **output:** Region list $\mathcal{R}$

where y indexes pixels. It is based on the observation that if all the label values for region R_i in M are altered after the convolution, R_i cannot encompass k_t. Thus, the scale of the region is smaller than t.

We present the scale estimation process in Algorithm 1. After obtaining regions whose scales are smaller than t, we merge each of them to its closest neighboring region in CIELUV color space. The merging process is shown in Algorithm 2.

3.2 Single-Layer Saliency Cues

For each layer we extract, saliency cues are applied to find important regions from the perspectives of color, position and size. We present two cues that are particularly useful.

3.2.1 Local contrast

Image regions contrasting their surroundings are general eye-catching [9]. We define the local contrast saliency cue for R_i in an image with a total of n regions as a weighed sum of color difference from other regions:

$$C_i = \sum_{j=1}^{n} w(R_j)\phi(i,j)\|\mathsf{c}_i - \mathsf{c}_j\|_2, \qquad (3)$$

where c_i and c_j are colors of regions R_i and R_j respectively. $w(R_j)$ counts the number of pixels in R_j. Regions with more pixels contribute higher local-contrast weights than those containing only a few pixels. $\phi(i,j)$ is set to $\exp\{-D(R_i,R_j)/\sigma^2\}$ controlling the spatial distance influence between two regions i and j, where $D(R_i,R_j)$ is a square of Euclidean distances between region centers of R_i and R_j. With the $\phi(i,j)$ term, close regions have larger impact than distant ones. Hence, Eq. (3) measures color contrast mainly to surroundings. Parameter σ^2 controls how large the neighborhood is. It is set to the product of $(0.2)^2$ and the particular scale threshold for the current layer. In the top layer, σ^2 is large, making all regions be compared in a near-global manner.

3.2.2 Location heuristic

Human attention favors central regions [41]. So pixels close to the image center could be good candidates, which have been exploited in [38], [29]. Our location heuristic is written as

$$H_i = \frac{1}{w(R_i)} \sum_{x_i \in R_i} \exp\{-\lambda\|x_i - x_c\|^2\}, \qquad (4)$$

where $\{x_0, x_1 \cdots\}$ is the set of pixel coordinates in region R_i, and x_c is the coordinate of the image center. H_i makes regions close to image center have large weights. λ is a parameter used when H_i is combined with C_i, expressed as

$$\bar{s}_i = C_i \cdot H_i. \qquad (5)$$

Since the local contrast and location cues have been normalized to range $[0, 1)$, their importance is balanced by λ, set to 9 in general. After computing $\bar{s}_i$ for all layers, we obtain initial saliency maps separately, as demonstrated in Fig. 6(b)-(d).

In what follows, we describe how Fig. 6(e) is obtained from the three single-layer saliency maps through our local consistent hierarchical inference. This strategy is updated from the one presented in our conference version paper [46] in both construction and optimization. It leads to improved performance.

3.3 Local Consistent Hierarchical Inference

Cue maps reveal saliency in different scales and could be quite different. At the bottom level, small regions are produced while top layers contain large-scale structures. Due to possible diversity, none of the single layer information is guaranteed to be perfect. It is also hard to determine which layer is the best by heuristics.

Multi-layer fusion by naively averaging all maps is not a good choice, considering possibly complex background and/or foreground. Note in our region

(a) Input (b) Cue map in $\mathcal{L}^1$ (c) Cue map in $\mathcal{L}^2$ (d) Cue map in $\mathcal{L}^3$ (e) Final saliency
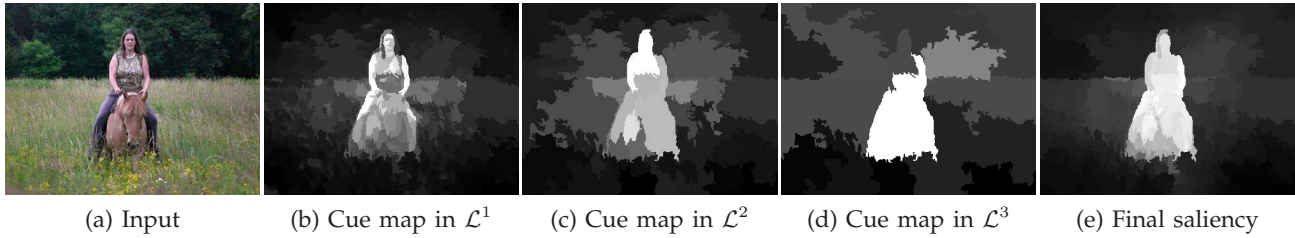
Fig. 6. Saliency cue maps in three layers and our final saliency map.

merging steps, a segment is guaranteed to be encompassed by the corresponding ones in upper levels. This makes a hierarchy of regions in different layers naturally form. An example is shown in Fig. 2(e). In the graph, the nodes in three layers correspond to regions from the three image layers. The connection between them in neighboring layers is due to the "belonging" relationship. For instance, the blue node j corresponds to the blue region in (d). It contains two segments in the lower level and thus introduces two children nodes, marked red and green respectively.

Without considering the connections between nodes in the same layer, the graph actually can be seen as a tree structure after adding a virtual node representing the entire image. The structure inspires a hierarchical inference model to take into account the influence of regions from neighboring layers, so that large-scale structures in upper layers can guide saliency assignment in lower layers.

In addition, if an object is narrow and small, pixels could be mistakenly merged to background regions, such as the first example shown in Fig. 7. In such cases, considering only the influence of corresponding regions in neighboring layers is insufficient. In our inference model, we count in a local consistency term between adjacent regions. Accordingly, in the graph shown in Fig. 2(e), connection between nodes in the same layer is built. We describe the process below.

3.3.1 Our Model

For a node corresponding to region i in layer $\mathcal{L}^k$, we define a saliency variable s_i^k. Set $\mathcal{S}$ contains all of them. We minimize the following energy function for the hierarchical inference

$$E(\mathcal{S}) = \sum_k \sum_i E_D(s_i^k) + \sum_k \sum_i \sum_{j, R_i^k \subseteq R_j^{k+1}} E_H(s_i^k, s_j^{k+1})$$

$$+ \sum_k \sum_i \sum_{j, R_j^k \in \mathcal{A}(R_i^k)} E_C(s_i^k, s_j^k). \tag{6}$$

The energy consists of three parts. Data term $E_D(s_i^k)$ is to gather separate saliency confidence, and hence is defined, for every node, as

$$E_D(s_i^k) = \beta^k ||s_i^k - \bar{s}_i^k||_2^2, \tag{7}$$

where β^k controls the layer confidence and $\bar{s}_i^k$ is the initial saliency value calculated in Eq. (5). The data term follows a common definition.

The hierarchy term $E_H(s_i^k, s_j^{k+1})$, building cross-layer linkages, enforces consistency between corresponding regions in different layers. This term is important in our inference model. It not only connects multi-layer information, but also enables reliable combination of saliency results among different scales. In detail, if R_i^k and R_j^{k+1} are corresponding in two layers, we must have $R_i^k \subseteq R_j^{k+1}$ based on our encompassment definition and the segment generation procedure. E_H is defined on them as

$$E_H(s_i^k, s_j^{k+1}) = \lambda^k ||s_i^k - s_j^{k+1}||_2^2, \tag{8}$$

where λ^k controls the strength of consistency between layers. The hierarchical term makes saliency assignment for corresponding regions in different levels similar, beneficial to effectively correcting single-layer saliency errors.

The last term is a local consistency term, which enforces intra-layer smoothness. It is used to make saliency assignment smooth between adjacent similar regions. Notation $\mathcal{A}(R_i^k)$ in Eq. (6) represents a set containing all adjacent regions of R_i^k in layer $\mathcal{L}^k$. If $R_j^k \in \mathcal{A}(R_i^k)$, the consistency penalty between regions R_i^k, R_j^k is expressed as

$$E_C(s_i^k, s_j^k) = \gamma^k w_{i,j}^k ||s_i^k - s_j^k||_2^2, \tag{9}$$

where γ^k determines the strength of consistency for each layer. $w_{i,j}^k$ is the influence between adjacent regions R_i^k and R_j^k. It should be large when the two regions are similar in color and structure. We define it as regional similarity in the CIELUV color space:

$$w_{i,j}^k = \exp\left\{-\frac{||c_i^k - c_j^k||^2}{\sigma_c}\right\}, \tag{10}$$

where c_i^k and c_j^k are mean colors of respective regions and σ_c is a normalization parameter. The intra-layer consistency brings local regions into consideration. Thus the inference is robust to hierarchical errors.

Our energy function including these three terms considers multi-layer saliency cues, making final results have less errors occurred in each single scale.

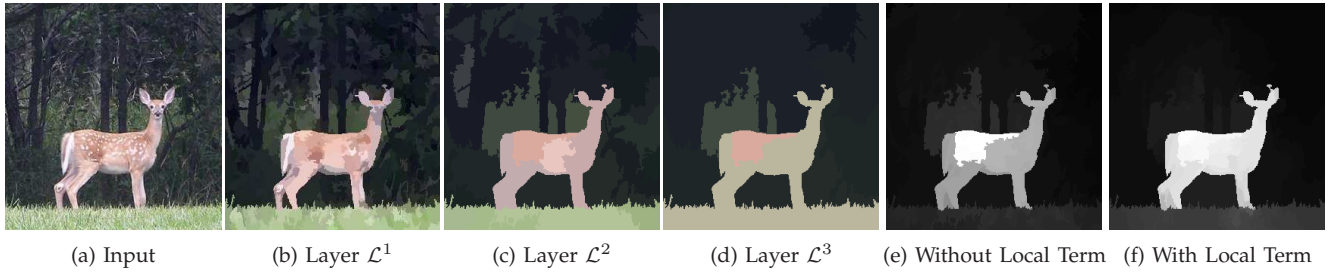| (a) Input | (b) Layer $\mathcal{L}^1$ | (c) Layer $\mathcal{L}^2$ | (d) Layer $\mathcal{L}^3$ | (e) Without Local Term | (f) With Local Term |

Fig. 7. Comparison of inference models with and without the local consistency term. Enforcing local connection makes the final saliency map [46] in (f) less affected by similar color in the background.

3.3.2 Optimization

Our objective function in Eq. (6) forms a simple hierarchical graph model. Since it contains loops inside each layer, we adopt common loopy belief propagation [32] for optimization. It starts from an initial set of belief propagation messages, and then iterates through each node by applying message passing until convergence.

The propagation scheme can be summarized as follows. Let $m_{i \to j}^{\tau}$ be the message passed from region R_i to an adjacent region R_j at the τ-th iteration (we omit layer indicator for simplicity here). At each iteration, if R_i and R_j are in the same layer, the message is

$$m_{i \to j}^{\tau}(s_j) = \min_{s_i} \left\{ E_D(s_i) + E_C(s_i, s_j) + \sum_{p \in \mathcal{N}(R_i) \backslash j} m_{p \to i}^{\tau-1}(s_i) \right\}, \quad (11)$$

where set $\mathcal{N}\{R_i\}$ contains connected region nodes of R_i, including inter- and intra-layer ones. If R_i and R_j are regions in different layers, the passed message is

$$m_{i \to j}^{\tau}(s_j) = \min_{s_i} \left\{ E_D(s_i) + E_H(s_i, s_j) + \sum_{p \in \mathcal{N}(R_i) \backslash j} m_{p \to i}^{\tau-1}(s_i) \right\}. \quad (12)$$

After message passing converges at T-th iteration, the optimal value of each saliency variable can be computed via minimizing its belief function, expressed as

$$(s_j)^* = \arg\min_{s_j} \left\{ E_D(s_j) + \sum_{i \in \mathcal{N}(R_j)} m_{i \to j}^{T}(s_j) \right\}. \quad (13)$$

Finally, we collect the saliency variables in layer $\mathcal{L}^1$ to compute the final saliency map. An example is shown in Fig. 6(e). Although saliency assignment in the original resolution is erroneous, our final saliency map correctly labels the woman and horse as salient. The background containing small-scale structures is with low and smooth saliency values, as expected from our model construction.

3.4 More Discussions

Relation to the Inference Model in [46] The inference model proposed in our conference version
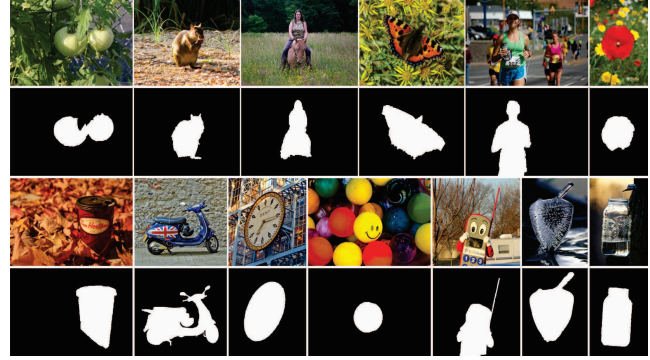

Fig. 8. Example images from Extended Complex Scene Saliency Dataset (ECSSD). The images in the first row contain complex structures either in the salient foreground or the non-salient background. The second row shows the corresponding objects marked by human.

paper [46] is a simplified version of Eq. (6), where γ^ks are set to 0 and linkage between adjacent regions in the same layer does not exist. It is written as

$$E(\mathcal{S}) = \sum_k \sum_i E_D(s_i^k) + \sum_k \sum_i \sum_{j, R_i^k \subseteq R_j^{k+1}} E_H(s_i^k, s_j^{k+1}), \quad (14)$$

forming a tree structure instead. This scheme enables simpler and more efficient optimization where belief propagation is still applicable. In this case, message passing is expressed in a single form by Eq. (12). It makes exact inference with global optimum achievable within two passes [25]. This type of inference is actually equivalent to applying a weighted average to all single-layer saliency cue maps, with optimally determined weight for each region.

The tree-structured inference model is capable to find commonly salient regions. However, without intra-layer propagation, narrow objects could be mismerged to background, as exemplified in Fig. 7. In the two examples, foreground and background pixels are mistakenly joined due to object similarity. Our new model counts in image layer information through the local consistency term and reduces such errors, as shown in (f). In Section 4, we show more quantitative and qualitative evaluation results.
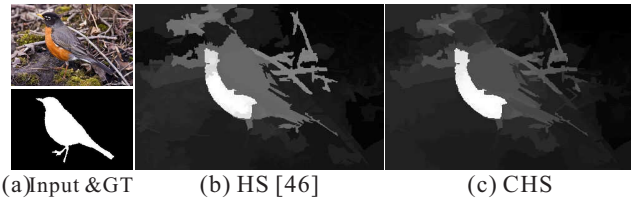
(a)Input &GT   (b) HS [46]   (c) CHS

Fig. 9. An example that our new CHS model does not outperform previous HS [46] model.

Our new model, adding a local consistency term based on regional color similarity, aims to correct saliency score due to mistaken merge. When the foreground and background have almost the same color, the new term may degrade the performance, as shown in Fig. 9. Under this extreme situation, our previous tree-structured inference model HS in [46] is more suitable.

Relation to other Hierarchical Models Similar ideas of combining information in a hierarchical structure were employed in this community. Wu *et al.* [45] proposed a tree-structured hierarchal model for short- and long-range motion perception. Ladicky *et al.* [26] developed an associative hierarchical random fields for semantic segmentation. In saliency related applications, Sun *et al.* [40] modeled eye movement for visual object attention via a hierarchical understanding of scene/object. The saliency output varies with gaze motion. A stream of bottom-up saliency models [20], [29], [1] explore information in different layers. These methods exploit independent information in different layers while our model connects saliency maps via a graphical model to produce final saliency scores. The interaction between layers enhances optimality of the system. Besides, our methods generally produce clearer boundaries than previous local-contrast approaches, as we do not employ downsampled images in different layers.

4 EXPERIMENTS

Our current un-optimized implementation takes on average 2.02s to process one image with resolution 400×300 in the benchmark data on a PC equipped with a 3.40GHz CPU and 8GB memory. The most time consuming part, taking 86% of the total time, is the local consistent hierarchical inference. Our implementation of this step is based on the Matlab package [36] for loopy belief propagation. Acceleration could be achieved by more efficient implementation. In our experiments, β_i is fixed as $\{0.5, 4, 2\}$ for $i = \{1, 2, 3\}$, and $\lambda_1 = \lambda_2 = 4$.

In what follows, we first introduce our Extended Complex Scene Saliency Dataset (ECSSD) for saliency

evaluation. Then we show both qualitative and quantitative experiment results of our method on this new and other benchmark datasets.

4.1 Extended Complex Scene Saliency Dataset

Although images from MSRA-1000 [2] have a large variety in their content, background structures are primarily simple and smooth. To represent the situations that natural images generally fall into, we extend our Complex Scene Saliency Dataset (CSSD) in [46] to a larger dataset (ECSSD) with 1000 images, which includes many semantically meaningful but structurally complex images for evaluation. The images are acquired from the internet and 5 helpers were asked to produce the ground truth masks individually.

We asked the helpers to label multiple salient objects if they think these objects exist. We evaluate the inter-subject label consistency by F-measure using four labels for ground-truth and the rest one for testing following the protocol of [28]. To build the ground-truth mask, we average the binary mask of the four labelers and set the threshold to 0.5. The averaged F-measure among the 5 cases in all ECSSD images is 0.9549, which indicates the inter-subject label is quite consistent although our dataset contains complex scenes. The final results are selected by majority vote, i.e., averaging 5 candidate masks and setting the threshold to 0.5. All images and ground truth masks in the dataset are publicly available.

Dataset Properties Images in our dataset fall into various categories. The examples shown in Fig. 8 include images containing natural objects like vegetables, flowers, mammals, insects, and human. There are also images of man-made objects, such as cups, vehicles, and clocks. For each example, we has its corresponding salient object mask created by human.

Backgrounds of many of these examples are not uniform but contain small-scale structures or are composed of several parts. Some of the salient objects marked by human also do not have a sharply clear boundary or obvious difference with the background. Natural intensity change due to illumination also exists. The fourth image in the upper row of Fig. 8 is a typical example because the background contains many flowers diversified in color and edge distributions; the foreground butterfly itself has high-contrast patterns. Considering only local center-surround contrast could regard all these high-contrast pixels as salient. Results by several recent methods are shown in Section 4.2.

In our image dataset, it is also noteworthy that multiple objects possibly exist in one image, while part of or all of them are regarded as salient decided by human. In the fourth example in third row of Fig. 8, several
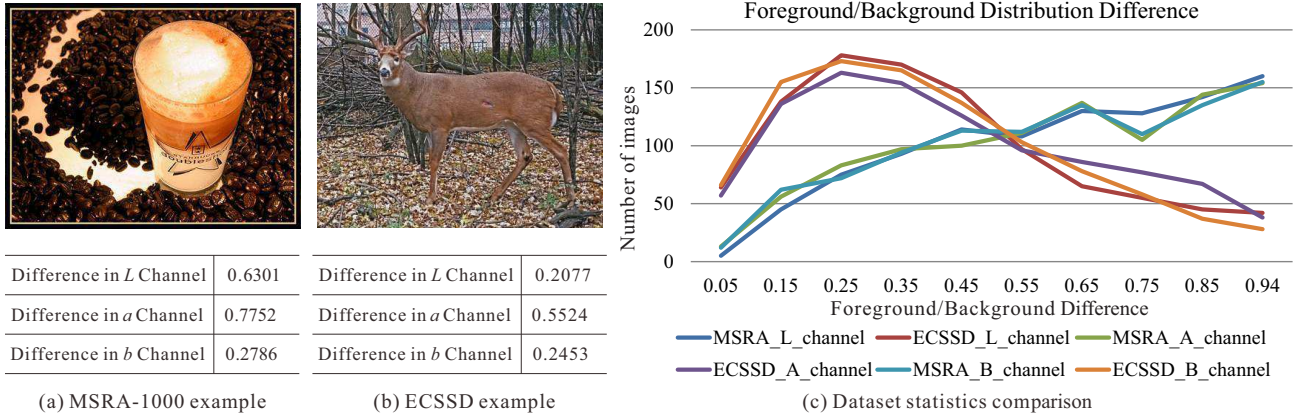
Difference in L Channel	0.6301
Difference in a Channel	0.7752
Difference in b Channel	0.2786

Difference in L Channel	0.2077
Difference in a Channel	0.5524
Difference in b Channel	0.2453

(a) MSRA-1000 example   (b) ECSSD example   (c) Dataset statistics comparison

Fig. 10. Dataset complexity comparison. (a) and (b) are from MSRA-1000 and ECSSD respectively. The latter is visually complex and also has a small foreground/background difference. Figure (c) shows the histogram of foreground/background difference on two datasets, evaluated on Ł, a, _ channels separately. It manifests that our dataset has more similar foreground/background pairs, thus becomes more difficult for saliency detection.

(a) Input   (b) GT   (c) FT [2]   (d) CA [14]   (e) HC [9]   (f) RC [9]   (g) RCC [8]   (h) HS [46]   (i) CHS
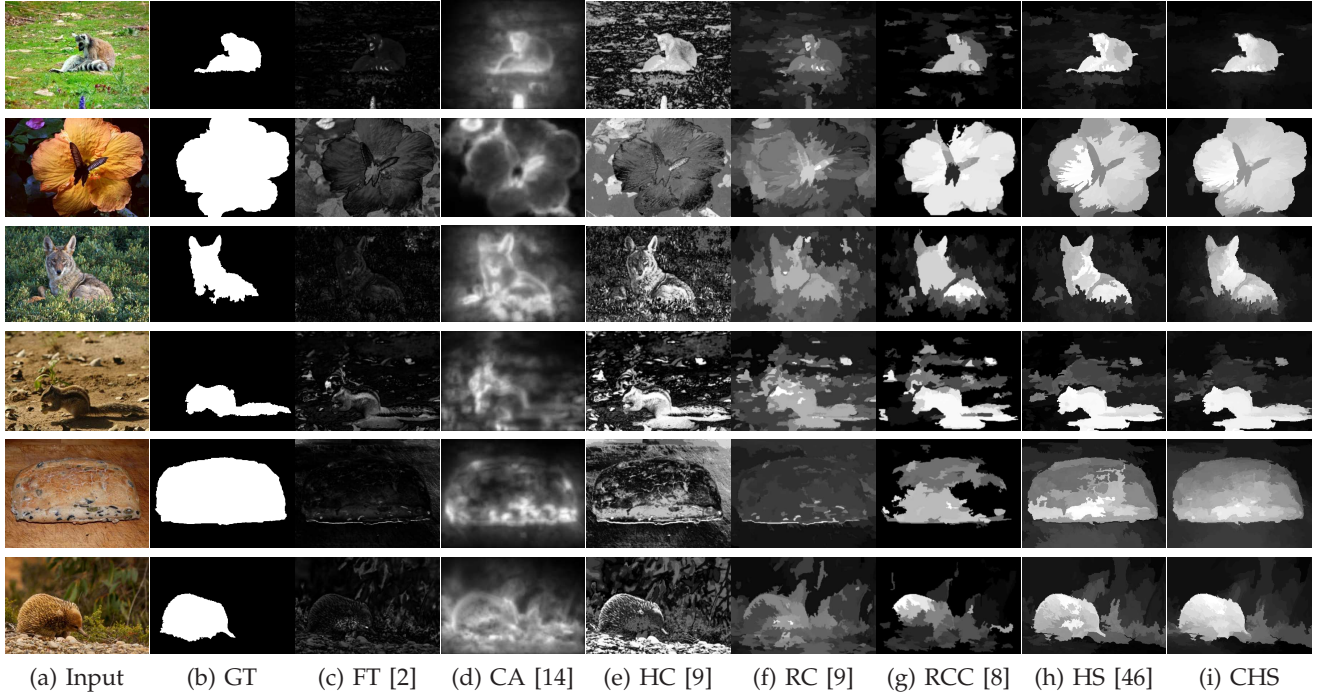
Fig. 11. Visual Comparisons on ECSSD.

balls with different colors are put together. Because the central ball has a smiling face, it is naturally more attractive. This is a challenging example for saliency detection.

In addition, this ECSSD dataset contains transparent objects with their color affected by background patterns, causing large ambiguity in detection. These salient objects nevertheless are easy to be determined by human. We hope, by including these difficult images, new definitions and solutions can be brought into the community in future for more successful and powerful saliency detection.

Complexity Evaluation We quantitatively evaluate the complexity of our dataset via the difference of foreground/background distribution in CIELab color space. Given the ground truth mask, we separate each image into foreground and background pixels. Then Chi-square distance is computed on the distributions of these two sets considering the Ł, a and _channels. Large difference values mean foreground and background can be easily separable, while a small difference increases the difficulty to distinguish foreground from background.

Two image examples are shown in Fig. 10(a) and (b) with their respective foreground/background distribution difference values. In Fig. 10(c), we plot the

histogram of the difference for all images included in MSRA-1000 and ECSSD respectively. It manifests that our dataset has many more images with low foreground/background difference compared to those in MSRA-1000. Put differently, our new dataset contains more complex images for saliency detection.

4.2 Evaluation on ECSSD

We evaluate our method on the ECSSD dataset and compare our results with those from several prior methods, including local schemes – IT [20], GB [16], AC [1] – and global schemes – LC [48], FT [2], CA [14], HC [9], RC[9], RCC [8], LR [38], SR[18]. The abbreviations are the same as those in [9], except for LR, which represents the low rank method of [38]. For IT, GB, AC, FT, CA, HC, RC, RCC, LR and SR, we run authors' codes. For LC, we use the implementation provided in [9]. We denote the tree-structured based method we proposed in [46] as HS, and our new method as CHS.

The visual comparison is given in Fig. 11. Our methods can handle complex foreground and background with different details, giving accurate and uniform saliency assignment. Compared with the tree-structured algorithm [46], our new local consistent hierarchical inference produces less turbulent saliency values among similar adjacent regions. More importantly, it is able to correct some foreground pixels that are mistakenly merged to the background. More results will be available on our project website.

In quantitative evaluation, we plot the precision-recall curves Fig. 12(a). Our experiment follows the setting in [2], [9], where saliency maps are binarized at each possible threshold within range $[0, 255]$. Our method achieves the top precision in almost the entire recall range $[0, 1]$. It is because combining saliency information from three scales makes the saliency estimation be considered both locally and globally. Only sufficiently salient objects through all scales are detected in this case. The non-salient background is then with low scores generally. Besides, adding the local consistency term improves performance by preserving consistent saliency between adjacent regions.

In many applications, high precision and high recall are required. We thus estimate the F-Measure [2] as

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}. \quad (15)$$

Thresholding is applied and β^2 is set to 0.3 as suggested in [2]. The F-measure is plotted in Fig. 12(b). Our methods have high F-scores compared to others in most range, indicating less sensitivity to picking a threshold in both versions.

We have further investigated the performance of *Mean absolute error* (MAE) following Perazzi *et al.*

[34]. MAE is a better representation for segmenting salient objects. Table 1 demonstrates that our HS and CHS outperform most existing methods by a large margin. Note that the result of HS is slightly better than CHS under this measure, since in our ECSSD dataset, there are many data with insignificant foreground/background difference. Recent work of RCC [8] performs similarly as ours. For relatively small thresholds, it enforces a boundary prior to produce clean background in the saliency map. The SaliencyCut step iteratively updates the saliency map to produce nearly binary maps. On complex background without confident saliency map initialization, the SaliencyCut step would remove the less confident region, such as the last two examples in Fig. 11.

4.3 MSRA-1000 [2], and 5000 [29] Dataset

We also test our method on the saliency datasets MSRA-1000 [2] and MSRA-5000 [29], [21] where MSRA-1000 is a subset of MSRA-5000, containing 1000 natural images. We show comparisons with the following ones, including local methods – IT [20], MZ [30], GB [16], AC [1], and global methods – LC [48], FT [2], CA [14] HC [9], RC [9], RCC [8], SF [34], SR [18]. For IT, GB, CA, RCC and SR, we run authors' codes. For AC, FT, HC, LC, MZ, RC and SF, we directly use author-provided saliency results. We omit result of HS here since the performance in this two datasets is rather close. Images of MSRA-1000 and MSRA-5000 are relatively more uniform; hence benefit of the local consistency term is not obvious.

Visual comparison is shown in Fig. 13. Follow previous settings, we also quantitatively compare our method with several others with their saliency maps available. The precision-recall curves for the MSRA-1000 and 5000 datasets are plotted in Fig. 14(a) and (c). The F-measure curves for the two datasets are plotted in Fig. 14(b) and (d). The MAE measures are listed in Tables 1. On these simpler datasets, all methods, including ours, perform much better. However, the advantage of our method is still clear.

4.4 Pascal-S Dataset [28]

We also compared our CHS and HS [46] on Pascal-S dataset [28] with SF [34], HC [9], IT [20], FT [2], RCC [8], and GBVS-CMPC [28]. Pascal-S dataset is newly proposed for benchmarking complexity images in saliency detection. Both Pascal-S and our ECSSD datasets contain complex images for saliency evaluation. The images in Pascal-S usually involve several different objects, such as a person in a bedroom with decorations and furniture. In our dataset, many similar foreground and background color/structure distributions make the images difficult for saliency
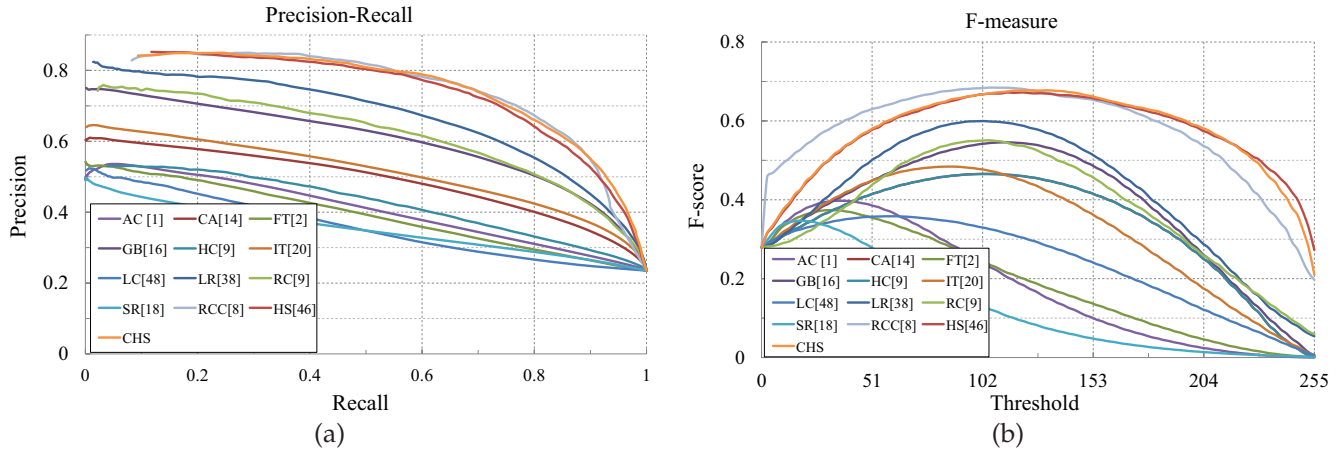
Fig. 12. Quantitative comparison on ECSSD.

	AC [1]	CA [14]	FT [2]	GB [16]	HC [9]	IT [20]	LC [48]	LR [38]	SR [18]	RC [9]	RCC [9]	HS [46]	CHS
ECSSD	0.2647	0.3100	0.2698	0.2821	0.3258	0.2900	0.2940	0.2669	0.2636	0.3005	0.1865	0.2244	0.2265
MSRA-1000	0.2102	0.2332	0.2053	0.2189	0.1774	0.1953	0.2187	0.1854	0.2149	0.2358	0.1062	0.1155	0.0961
MSRA-5000	0.2280	0.2503	0.2298	0.2433	0.2391	0.2475	0.2447	0.2152	0.2251	0.2638	0.1399	0.1528	0.1499

TABLE 1
Quantitative comparison for MAE on ECSSD, MSRA-1000, and MSRA-5000 datasets.

(a) Input   (b) GT   (c) MZ [30]   (d) LC [48]   (e) GB [16]   (f) RC [9]   (g) SF [34]   (h) RCC [8]   (i) CHS
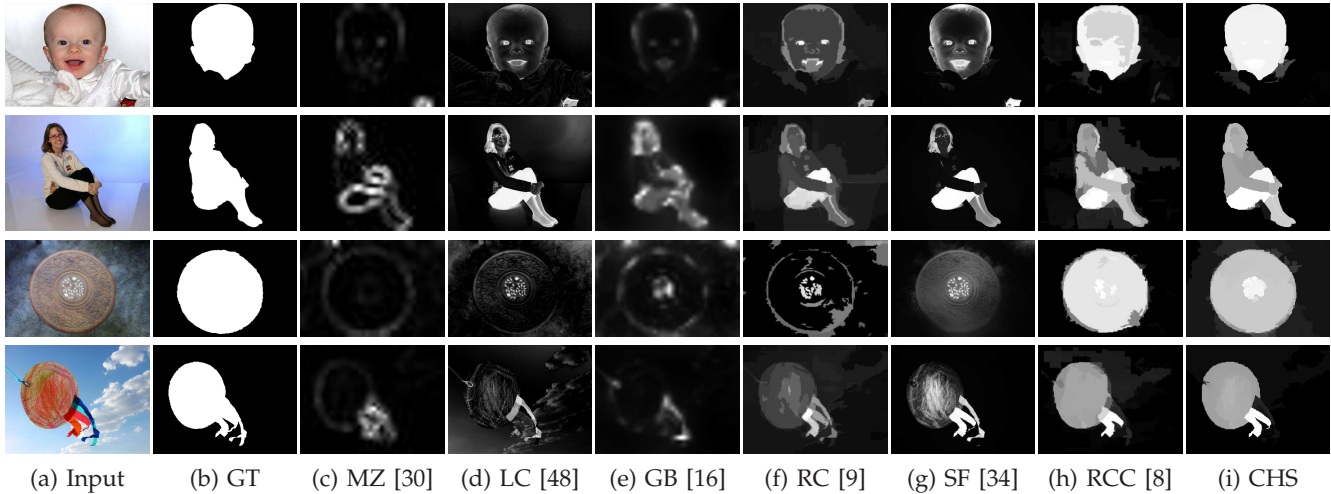
Fig. 13. Visual comparison on MSRA-1000 [2].

detection. Most of our data contain the only salient object as ground truth without ambiguity (see examples in Fig. 11).

The precision-recall comparison is shown in Fig. 15. The success of GBVS-CMPC [28] is due to the fact that the CPMC segmentation algorithm it used has already produced decent object-level proposals. Salient object detection is achieved by assigning a uniform score map to those confident objects. GBVS-CMPC shows that the eye-fixation together with an excellent segmentation approach is a promising direction.

4.5 Comparison with Single-Layer

Our hierarchical framework utilizes information from multiple image layers, gaining special benefit. Single-layer saliency computation does not work similarly well. To validate it, we take $\bar{s}_i$ in Eq. (5) in different layers as well as the average of them as the saliency values. We evaluate how they work respectively when applied to our ECSSD image data.

We compare all single layer results, averaged result, result by tree-structured inference in [46] and result by our local consistent hierarchical inference, denoted as Layer1, Layer2, Layer3, Average, HS, and CHS respectively. For each of them, we take the threshold that maximizes F-measure, and plot the correspond-
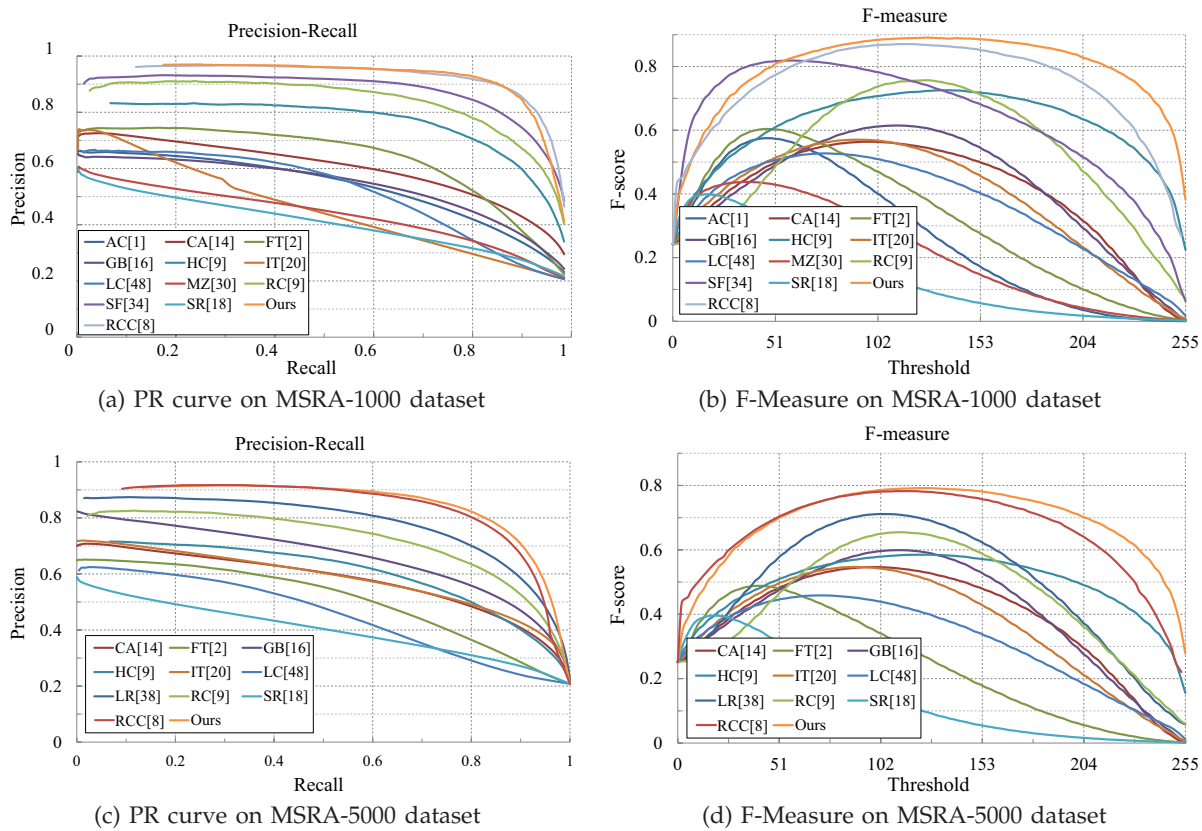
(a) PR curve on MSRA-1000 dataset

(b) F-Measure on MSRA-1000 dataset

(c) PR curve on MSRA-5000 dataset

(d) F-Measure on MSRA-5000 dataset

Fig. 14. Precision-recall / F-measure curve on MSRA-1000 and MSRA-5000 datasets.
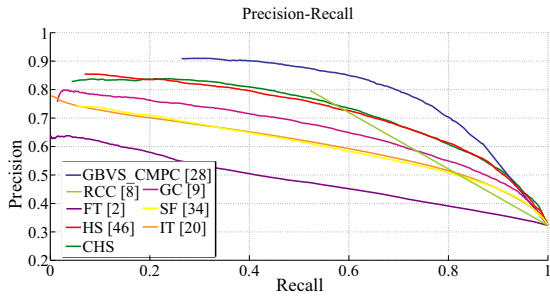
Fig. 15. Precision-recall curve on Pascal-S dataset.

	Layer1	Layer2	Layer3	Average	HS	CHS
Precision	0.7278	0.7124	0.7275	0.7468	0.7360	0.7639
Recall	0.6489	0.6773	0.6236	0.6584	0.6832	0.6564
F-measure	0.6489	0.6437	0.6440	0.6739	0.6721	0.6776

TABLE 2
Performance of Single-layer vs. multi-layer. For each
method, we take the threshold that corresponds to the
highest F-measure, and list the precision, recall and
F-measure together.

	5	9	13	17	21	25	29	33
Traditional	0.682	0.680	0.679	0.694	0.695	0.703	0.700	0.705
Ours	0.685	0.696	0.712	0.738	0.763	0.760	0.805	0.806

TABLE 3
Performance of traditional scale measure vs. our scale
measure for F-measure under 7 different scales.

ing precision, recall, and F-measure in Table 2.

Results from all single layers are close. But the performance decreases. The reason is that, as more small-scale structures are removed, the extracted image layers are prone to segmentation errors especially for the structurally complex images in our dataset. On the other hand, large-scale image layers benefit large-scale result representation. Compared with naive averaging of all layers, our inference algorithm optimally aggregates confident saliency values from these layers, surely yielding better performance. By enforcing smoothness locally in our new inference model, CHS also produces better results compared to the simpler HS implementation.

4.6 Region Scale and Layer Number Evaluation

To evaluate the effectiveness of our new scale measure presented in Section 3.1, we compare our results with those produced using the traditional scale measure, i.e., number of pixels in the region. We replace the scale measure by counting pixel number and set scale thresholds $\{5, 9, 13, 17, 21, 25, 29, 33\}$ for our measure and squares of these values for the traditional one.

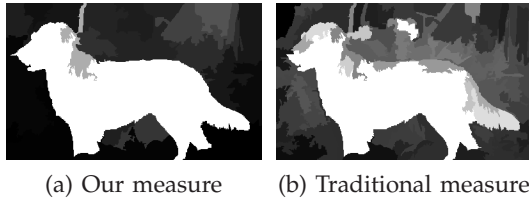For images containing text- or curve-like high contrast

(a) Our measure   (b) Traditional measure

Fig. 16. Scale measure comparison illustration.

nLayers	2	3	4	5
F-measure	0.6736	0.6776	0.6740	0.6710

TABLE 4

The maximun F-measure for different layer number.

regions that should not be classified as salient alone, our method performs much better. The resulting F-measure scores for the representative images from ECSSD are listed in Table 3, indicating that our new region scale measure is effective. Fig. 16 shows an image result comparison. Uniformity-enforced scale measure profits general saliency detection and can remove errors caused by detecting many narrow regions in the fine level.

We also evaluate how the number of layers affects our result. We experiment using different number of layers and adjust the single layer scale for the best performance on ECSSD. F-measure is reported in Table 4. The three-layer case produces the best result. Two layers cannot take a comparable advantage of scale variation. With more than three layers, more errors could be introduced in large scales. Also the computational time increases accordingly.

5 CONCLUDING AND FUTURE WORK

We have tackled a fundamental problem that small-scale structures would adversely affect salient detection. This problem is ubiquitous in natural images due to common texture. In order to obtain a uniformly high-response saliency map, we propose a hierarchical framework that infers importance values from three image layers in different scales. Our proposed method achieves high performance and broadens the feasibility to apply saliency detection to more applications handling different natural images. The future work includes incorporating object segmentation clues, and applying the hierarchical insight to salient eye-fixation and objectness methods.

ACKNOWLEDGEMENTS

REFERENCES

[1] R. Achanta, F. J. Estrada, P. Wils, and S. Süsstrunk. Salient region detection and segmentation. In *ICVS*, pages 66–75, 2008.

[2] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604, 2009.

[3] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *TPAMI*, 34(11):2189–2202, 2012.

[4] A. Borji and L. Itti. Exploiting local and global patch rarities for saliency detection. In *CVPR*, pages 478–485, 2012.

[5] A. Borji, D. N. Sihite, and L. Itti. Salient object detection: A benchmark. In *ECCV*, pages 414–429, 2012.

[6] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai. Fusing generic objectness and visual saliency for salient object detection. In *ICCV*, pages 914–921, 2011.

[7] K.-Y. Chang, T.-L. Liu, and S.-H. Lai. From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model. In *CVPR*, pages 2129–2136, 2011.

[8] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu. Global contrast based salient region detection. *TPAMI*, 37(3):569–582, 2015.

[9] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Global contrast based salient region detection. In *CVPR*, pages 409–416, 2011.

[10] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. In *CVPR*, 2014.

[11] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *TPAMI*, 24(5):603–619, 2002.

[12] I. Endres and D. Hoiem. Category independent object proposals. In *ECCV*, pages 575–588. 2010.

[13] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004.

[14] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *CVPR*, pages 2376–2383, 2010.

[15] R. Gonzalez and R. Woods. *Digital image processing*. Prentice Hall Press, ISBN 0-201-18075-8, 2002.

[16] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, pages 545–552, 2006.

[17] P. Hiremath and J. Pujari. Content based image retrieval using color boosted salient points and shape features of an image. *International Journal of Image Processing*, 2(1):10–17, 2008.

[18] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *CVPR*, 2007.

[19] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, 2000.

[20] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *TPAMI*, 20(11):1254–1259, 1998.

[21] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient object detection: A discriminative regional feature integration approach. In *CVPR*, pages 2083–2090, 2013.

[22] T. Judd, K. A. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *ICCV*, pages 2106–2113, 2009.

[23] P. Khuwuthyakorn, A. Robles-Kelly, and J. Zhou. Object of interest detection by saliency learning. In *ECCV (2)*, pages 636–649, 2010.

[24] D. A. Klein and S. Frintrop. Center-surround divergence of feature statistics for salient object detection. In *ICCV*, 2011.

[25] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.

[26] L. Ladicky, C. Russell, P. Kohli, and P. Torr. Associative hierarchical random fields. Technical report, 2013.

[27] T. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *JOSA A*, 20(7):1434–1448, 2003.

[28] Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille. The secrets of salient object segmentation. In *CVPR*, 2014.

[29] T. Liu, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. In *CVPR*, 2007.

[30] Y.-F. Ma and H. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia*, pages 374–381, 2003.

[31] E. Macaluso, C. Frith, and J. Driver. Directing attention to locations and to sensory modalities: Multiple levels of selective processing revealed with pet. *Cerebral Cortex*, 12(4):357–368, 2002.

[32] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *UAI*, pages 467–475, 1999.

[33] S. Palmer. *Vision science: Photons to phenomenology*, volume 1. MIT press Cambridge, MA, 1999.

[34] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 2012.

[35] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi. Statistical textural distinctiveness for salient region detection in natural images. In *CVPR*, 2013.

[36] M. Schmidt. Ugm: Matlab code for undirected graphical models. http://www.di.ens.fr/ mschmidt/Software/UGM.html.

[37] G. Sharma, F. Jurie, and C. Schmid. Discriminative spatial saliency for image classification. In *CVPR*, pages 3506–3513, 2012.

[38] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *CVPR*, 2012.

[39] P. Siva, C. Russell, T. Xiang, and L. Agapito. Looking beyond the image: Unsupervised learning for object saliency and detection. In *CVPR*, 2013.

[40] Y. Sun, R. Fisher, F. Wang, and H. M. Gomes. A computer vision model for visual-object-based attention and eye movements. *CVIU*, 112(2):126–142, 2008.

[41] B. Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14), 2007.

[42] A. Toet. Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. *TPAMI*, 33(11):2131–2146, 2011.

[43] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch. Attentional selection for object recognition a gentle way. In *Biologically Motivated Computer Vision*, pages 251–267, 2002.

[44] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *ECCV*, 2012.

[45] S. Wu, X. He, H. Lu, and A. L. Yuille. A unified model of short-range and long-range motion perception. In *NIPS*, pages 2478–2486, 2010.

[46] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *CVPR*, 2013.

[47] C. Yang, L. Zhang, H. Li, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, 2013.

[48] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*, pages 815–824, 2006.

Jianping Shi received the BS degree in computer science and engineering from Zhejiang University, China in 2011. She is now pursuing her PhD degree in the Chinese University of Hong Kong. She received the Hong Kong PhD Fellowship and Microsoft Research Asia Fellowship Award in 2011 and 2013 respectively. Her research interests include in computer vision and machine learning. She is a student member of the IEEE.

Qiong Yan received her Ph.D. degree in computer science and engineering from Chinese University of Hong Kong in 2013 and the Bachelor's degree in computer science and technology from University of Science and Technology of China in 2009. She is now a researcher in Image and Visual Computing Lab in Lenovo. Her research interest is in saliency detection, image filtering and image enhancement. She is a member of the IEEE.

Li Xu received the BS and MS degrees in computer science and engineering from Shanghai JiaoTong University (SJTU) in 2004 and 2007, respectively, and the PhD degree in computer science and engineering from the Chinese University of Hong Kong (CUHK) in 2010. He joined Lenovo R&T Hong Kong in Aug 2013, where he leads the imaging & sensing group in the Image & Visual Computing (IVC) Lab. Li received the Microsoft Research Asia Fellowship Award in 2008 and the best paper award of NPAR 2012. His major research areas include motion estimation, motion deblurring, image/video analysis and enhancement. He is a member of the IEEE.
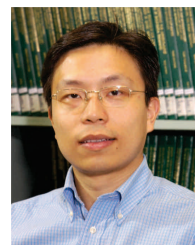
Jiaya Jia received the PhD degree in computer science from the Hong Kong University of Science and Technology in 2004 and is currently a professor in Department of Computer Science and Engineering, The Chinese University of Hong Kong (CUHK). He heads the research group focusing on computational photography, machine learning, practical optimization, and low- and high-level computer vision. He currently serves as an associate editor for the IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) and served as an area chair for ICCV 2011 and ICCV 2013. He was on the technical paper program committees of SIGGRAPH Asia, ICCP, and 3DV, and co-chaired the Workshop on Interactive Computer Vision, in conjunction with ICCV 2007. He received the Young Researcher Award 2008 and Research Excellence Award 2009 from CUHK. He is a senior member of the IEEE.