Deep Salience: Visual Salience Modeling via Deep Belief Propagation

Richard Jiang
Automatic Control and System Engineering
The University of Sheffield
Sheffield, United Kingdom
E-mail: richardjiang@acm.org

Danny Crookes
ECIT Institute
Queen's University Belfast
Belfast, United Kingdom
E-mail: d.crookes@qub.ac.uk

Abstract

Visual salience is an intriguing phenomenon observed in biological neural systems. Numerous attempts have been made to model visual salience mathematically using various feature contrasts, either locally or globally. However, these algorithmic models tend to ignore the problem's biological solutions, in which visual salience appears to arise during the propagation of visual stimuli along the visual cortex. In this paper, inspired by the conjecture that salience arises from deep propagation along the visual cortex, we present a *Deep Salience* model where a multi-layer model based on successive Markov random fields (sMRF) is proposed to analyze the input image successively through its deep belief propagation. As a result, the foreground object can be automatically separated from the background in a fully unsupervised way. Experimental evaluation on the benchmark dataset validated that our Deep Salience model can consistently outperform eleven state-of-the-art salience models, yielding the higher rates in the precision-recall tests and attaining the best *F*-measure and mean-square error in the experiments.

Introduction

Automated detection of visual objects in images and videos is a subject of primary interest because of its wide application in image/video indexing, content-aware editing, medical image analysis, intelligent computer-human interface, robotic vision, and visual surveillance. Researchers in artificial intelligence and computer vision have successfully developed a number of methods for object detection, such as AdaBoost face detection (Viola and Jones 2001), SVM-based human detection (Vedalid *et al* 2009; Dalal and Triggs 2005), and min-cut object segmentation (Rother *et al* 2004). These approaches usually depend on training on predefined datasets, or on user input such as scribbles or trimaps. However, when no prior knowledge of image content is available, unsupervised object detection is a hard problem, and it has attracted considerable interest from the research community.

The past decade has seen consistent progress towards unsupervised image segmentation and object detection. A widely-adopted approach is to consider an image principally as a set of hierarchical contours (Arbelaez *et al* 2011). This assumes that semantic content and the objects usually correspond to specific image structures. Recent research (Farabet *et al* 2013; Kohli *et al* 2013) has also suggested that this hierarchical view of image content may be correlated to the deep learning of image structures (Hinton *et al* 2006). In a development of this approach, we propose in this paper a *Deep Salience* model based on successive Markov Random Fields (sMRF) for unsupervised object detection. Our work is conceptually related to the recent pioneering work on hierarchical image analysis (Farabet *et al* 2013; Kohli *et al* 2013).

Unsupervised object detection usually leads to the topic of visual salience, which stems from psychological research on biological visual perception (Koch and Ullman 1985; Itti and Koch 2001). The earliest bio-inspired computational salience model was proposed by Koch and Ullman (1985), where the contrast between visual stimuli (pixels) was considered as the origin of salience awareness. A number of publications (Itti *et al* 1998; Ma and Zhang 2003; Harel *et al* 2006; Hou and Zhang 2007; Judd 2009) have followed this roadmap to develop their salience models using a variety of features. These methods are usually based on local contrast and tend to produce higher salience values near edges instead of uniformly highlighting salient objects. Cheng *et al* (2011) categorized these approaches as *local* approaches. Recent efforts have been made towards using *global* contrasts, where pixels or regions were evaluated with respect to the entire image. Achanta *et al* (2009) proposed a frequency tuned method that defines pixel salience using region-averaged contrast. Goferman *et al* (2012) used block-based global contrast. Cheng *et al* (2011) extend Achanta's method to region-based salience estimation. Perazzi *et al* (2012) further extend this region-contrast approach by leveraging superpixels. Jiang and Crookes (2012) used mutual information (MI) evaluation with a center-surround *a priori* map for global salience estimation. Peng *et al* (2013) introduced the low-rank matrix computation for salience modeling. In summary, both local and global methods have been based on modeling salience using various visual contrast definitions with various features based on pixels, blocks or regions.

From a biological viewpoint, we consider the conjecture that human visual salience is a consequence of the deep

propagation of visual stimuli along the human visual cortex (Ekstorm *et al* 2008). Hence, it is a complex procedure rather than a simple visual comparison using color contrast. To emulate this biological process, we present a *Deep Salience* model using our proposed successive Markov Random Fields (sMRF) model.

In our sMRF model, a fuzzy graph is introduced with a Markov Random Field to allow each node to maximize its belief by solving the fuzziness of edges. When applied to a cascaded multi-scale image pyramid, the pyramid structure of multiple-layer MRFs forms successive MRFs (sMRFs) which processes the input visual signals via its deep belief propagation along layers, while a belief maximization procedure is applied to detect salient objects automatically from its iterative convergence. Fig.1 shows an example of applying our *Deep Salience* method to a challenging image.

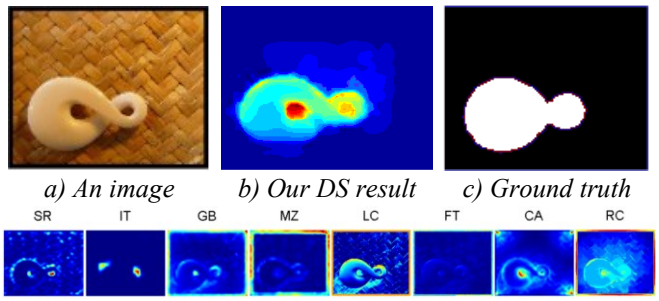
Successive Image Analysis

In image structure analysis, hierarchical analysis has been effective in tackling various computer vision tasks. Lowe (2004) proposed the use of a pyramid computation of DoG for multi-scale feature point detection. Sun and Pfister (2013) applied a pyramid-based coarse-to-fine approach to optical flow and achieved improved accuracy. In these methods, images were processed at different scales and the computation was carried out hierarchically. In the recent research (Arbelaez *et al* 2011; Farabet *et al* 2013; Kohli *et al* 2013), images have been considered as a collection of hierarchical contours. While hierarchical methods have been seen very successful in these challenging vision tasks, in this paper we apply this strategy to our visual salience analysis for unsupervised object detection.

In our method, the first step is to compute a pyramid of hierarchical images. Usually this can be done by using discrete wavelet transform (DWT). Given an input image, the 2D DWT computation will result in four components, namely LL, LH, HL and HH, as shown in Fig.2-a. In our approach, we keep only the LL (low-pass) component of each level. Consequently, we obtain a hierarchical pyramid of resized images at different resolutions, as shown in Fig.2-b. Our *Deep Salience* method will then take these images as input, and perform hierarchical analysis successively from the top of the pyramid downwards. The proposed coarse-to-fine process is akin to an emulation of the human visual system where the input visual signals are propagated along the visual cortex from eyes to brain (Koch and Ullman 1985; Ekstorm *et al* 2008).

Maximum Belief Propagation in MRF

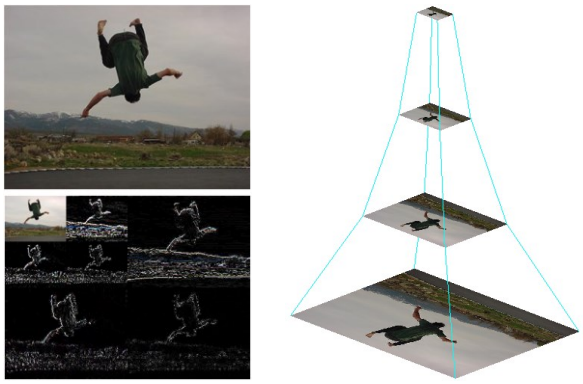
Markov random fields have been a powerful tool for image analysis. Given an undirected graph $G = (V, E)$ and observations $X=\{x_k\}$ with $k \in V$ indexed by V, a set of random variables $Y=\{y_k\}$ forms a conditional random field



a) An image *b) Our DS result* *c) Ground truth*

d) Salience maps from several state-of-the-art methods

Fig.1. A simple example that challenges most state-of-the-art methods. a) The image; b) Our Deep Salience(DS) map; c) Ground truth; d) Results by several state-of-the-art salience methods, SR (Hou & Zhang 2007), IT (Itti et al in 1998), GB (Harel et al 2006), MZ (Ma & Zhang 2003), LC (Zhai & Shah 2006), FT (Achanta et al 2009), CA (Goferman et al 2010), and RC (Cheng et al 2011). Here the 'jet' color map is applied to visualize the results.



a) Successive DWT *b) Hierarchical pyramid*

Fig.2. Hierarchical image analysis with wavelet pyramid.

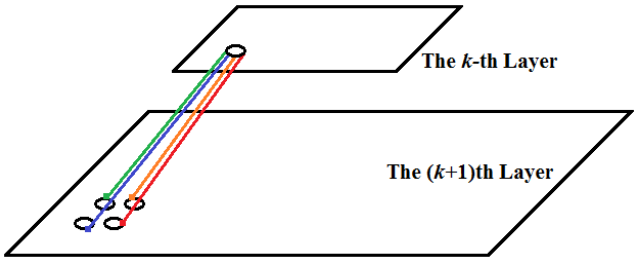
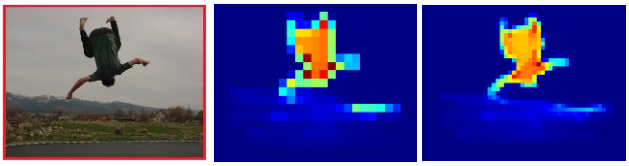


Fig.3 The proposed successive Markov random fields allow multi-layer deep belief propagation downwards along directed edges.

with respect to G when they satisfy the local Markov properties that a variable is conditionally independent of any variables that are not its neighbors. Hence, global belief propagation over a CRF can then be simplified as local belief propagation over cliques.

Our work starts from the pair-wise MRF model proposed by Jegelka and Bilmes (2011). Let V be the set of pixels in the image X we want to process, and let the set E contain all pairs of neighboring pixels that are edges in the MRF. The



a) Sample image *b) Level 4* *c) Level 3*

d) Level 2 *e) Level 1* *f) Level 0*

Fig.4. Deep Salience: Successive salience analysis via deep belief propagation over sMRF.

Images *QRF (0.1sec)* *MRF (3.5sec)*

Fig.5. DS modeling with sMRF using our QRF or standard MRF at each layer. QRF-based DS takes only 0.1s and highlights the foreground object, while typical MRF based DS needs longer time (3.5s) and blurs the region borders.

Images *One-layer MRF* *sMRF*

Fig.6. sMRF vs one-layer MRF: sMRF helps overcome the texture problem by its deep belief propagation.

label of each pixel in V is given by a binary random variable y_i whose labels (0/1) denote the label "foreground"/"background".

Considering a clique of two vertices i and j, if the confidence on the i-th vertex is known as P_i, the belief propagation along the clique edge can be inferred as a MAP (maximum a posteriori) formula:

$$P_i = P_j \theta_{ij} \qquad (1)$$

θ_{ij} denotes the conservative probability from the j^{th} vertex to the i^{th} vertex, which can be simply formulated with the feature difference between the i^{th} and j^{th} pixels:

$$\theta_{ij} = \psi\left(\left\|x_i - x_j\right\|\right) \qquad (2)$$

Here, $\psi(\cdot)$ denotes a penalty function when the belief is propagated across two color-different pixels. For example, given a pixel k_0 with a confidence P_0, a neighbor k_1 with exactly the same color should have a confidence $P_1=P_0$ with $\theta_{ij}=0$.

A node in a MRF may have multiple edges. Typical MRF methods using mean field continuum approximation (Yedidia *et al* 2005) may easily bleach the edges in an image and allow the belief propagation to cross region borders. To address this problem, we introduce a maximum belief propagation algorithm. We consider all edges of a node as a potential fuzzy edge in a fuzzy graph (Kosko 1986; Blue and Bush 2002; Salzenstein and Collet 2006) $\widetilde{G} = \left(V, \widetilde{E}\right)$. Hence, the pixel may have multiple choices of likelihood P_i^j inferred from different edges, and the fuzziness can be described by:

$$Fi = \sum_{j \in C} f_{ij} P_i^j = \sum_{j \in C} f_{ij} P_j \theta_{ij} \qquad (3)$$

Here, f_{ii} stands for the fuzziness of the j-th fuzzy edge. To maximize the belief in propagation, we apply a winner-takes-all strategy in the local iteration over random fields,

$$f_{ij} = \begin{cases} 1, \text{when } j = \arg\max_j P_i^j \\ 0, \text{otherwise} \end{cases} \qquad (4)$$

In the above winner-takes-all strategy, a node always favors the edge with the maximum confidence and dynamically sets other fuzzy edges as disconnected.

A major advantage of the above maximum belief propagation (MBP) algorithm is to guarantee that belief propagation in MRF will not blur the region borders in an input image. Besides, by disconnecting some edges dynamically, MBP inference problem is simplified into a local search of the winner in the fuzzy set of neighbors. In an analogy, the above MBP is more like a quantum mechanism (Birkhoff and Neumann 1936) that allows each node to switch between multiple quantum energy levels, while each fuzzy edge stands for a specific quantum energy level. For our convenience, we refer to our above simple implementation of MBP random fields as *Quantum Random Fields* (QRF) in this paper. Fig.5 shows the different belief propagation results in our *Deep Salience* model using typical MRF and our QRF. It is shown that QRF can easily converge on the salient object quickly.

Deep Belief Propagation via Successive MRF

Successive Markov Random Field

In research in unsupervised image segmentation and object detection, a common approach is to consider an image as a set of hierarchical contours (Arbelaez *et al* 2011). Recent

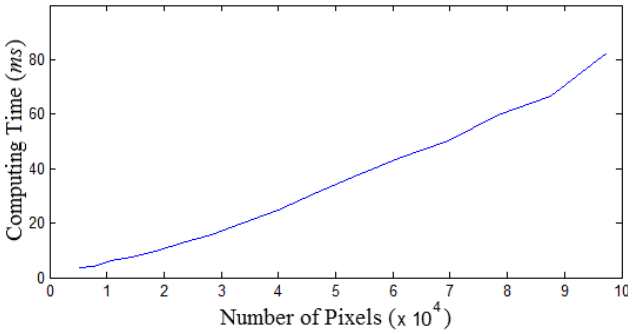


Fig.7. Computing time versus the number of pixels N.

Table I. Convergence time vs the number of layers (ms)

Layers	1 Layer	2 Layer	3 Layer	4 Layer	5 Layer
1st	210	103	78	61	12
2nd	--	29	16	14	4
3rd	--	--	6	3	3
4th	--	--	--	2	2
5th	--	--	--	--	2
Total	210	132	100	80	23

research (Farabet *et al* 2013; Kohli *et al* 2013) has also suggested that this hierarchical view of image content may be correlated to the deep learning of image structures.

As mentioned above, a pyramid of hierarchical images can be obtained from an input image using DWT, as shown in Fig.2. Each image in the pyramid can be mapped as an observation for each layer of successive MRF (sMRF). Fig.3 shows the concept of sMRF, where fuzzy edges not only present in one layer of random fields, but also connect a layer to the previous and next layers. Hence, the belief propagation is carried out down the pyramid of random fields. For a node in our successive QRF (sQRF), its inter-layer edges can be added into the search of its maximum belief,

$$j_m = \arg\max_{j,\ell}\{P_i^j, P_i^\ell\} \qquad (5)$$

Here P_i^ℓ is the estimated belief propagated from the nodes in other layers. In this work, we assume that the inter-layer connection is unidirectional so that the belief can only propagate from the upper layer to its lower layer. This implies an inheritance of the confidence map hierarchically down the pyramid from coarse to fine.

Starting from the top level of sMRF, we carry out the maximum belief propagation process (in Eq.(5)) iteratively over each layer of our *quantum random field* to solve its fuzziness. The computation on the top image (the one resized to $1/2^k$) gives the result as the k^{th} level belief map G_k. Then the belief map is propagated to the next layer down the pyramid.

Fig.4 shows an example of sMRF. Fig.4-a shows the sample image where, similar to Judd's work (2009), we assume that most pixels on the borders are taken as background (P_B=1). With this priori, the belief is then propagated based on the coarse observation from the top level, and successively downward to the lower sMRF layers. It can be seen that the resolution of the confidence maps increases and the details are gradually added via deep belief propagation. In our experiment, we use SF method (Perazzi *et al* 2012) to initialize the border and keep non-salient pixels as initial background.

The benefits of sMRF over single-layer MRF stem from its successive analysis of hierarchical image structures. It focuses on global structures at its initial coarse level and gradually focuses on local details in the subsequent levels. Therefore, it can allow belief propagation to go across texture regions easily. Fig.6 shows several examples comparing sMRF against single-layer MRF. We can see that the belief propagation in single-layer MRF gets stuck in texture-like regions in the background, while sMRF can overcome these local textures and highlight the salient foreground objects successfully in these test images.

Computing Complexity

Usually, the computing complexity of MRF is not readily predictable. However in our sQRF, the problem can become much easier. While the belief propagation is simplified as a local maximum problem in Eq.(5), the computing time is proportional to the range for a belief to propagate from the source to the pixel, which is related to the number of pixels. As shown in Fig.7, our experiment validated that the convergence time of sQRF is nearly proportional to the number of pixels, namely O(N).

It is noted that sMRF needs to converge on multiple layers. However, the number of nodes has been reduced to $1/2^k$, and the search space at the top level of the pyramid is then drastically reduced to $(1/2^k)^2$ as well. While the sMRF iteration moves downwards, the pre-converged confidence map is inherited by the next level and hence it can be expected to help reduce the total convergence time. Table I gives the experimental results of the computation times at different layers. The computing time was measured for a MATLAB solution. We can see that the single-level MRF took the longest time to converge (210 ms), while 5-level sMRF needs only 23 ms in total for all five levels. From this comparison we can see that successive MRF not only helps overcome texture background, but also helps reduce the computing time.

Experiments

We tested our method on the widely used object dataset --- the EPFL object dataset (Achanta *et al* 2009), which is publicly available and the ground truth of foreground objects is provided as binary masks. Most state-of-the-art methods have been reported with their benchmark results on this dataset. The EPFL dataset has the same images as MSRA-1000 except that it takes segmented objects as ground truth instead of gaze points from psychological

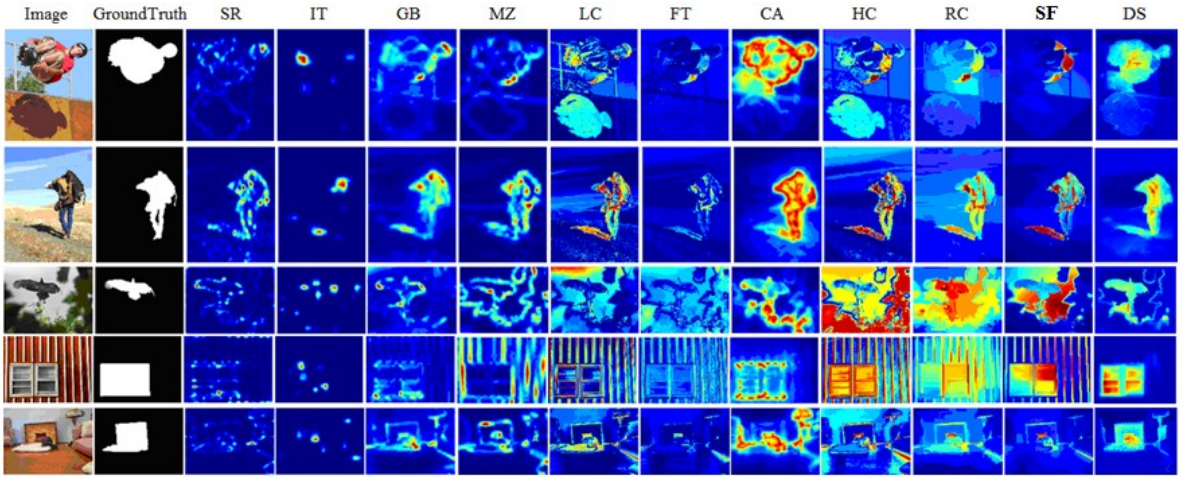
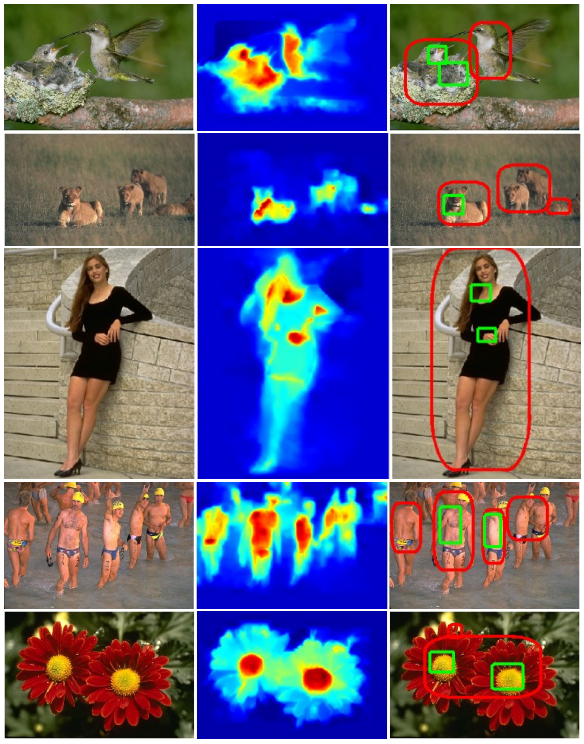


Fig.8. Visual comparison of salience maps. From left to right columns: original images; ground truth; salience maps from 1) SR; 2) IT; 3) GB; 4) MZ; 5) LC; 6) FT; 7) CA; 8) HC; 9) RC; 10) SF; and our Deep Salience (DS) maps. It can be seen that our DS model can robustly highlight salient objects in images where other algorithms failed.



a) Input images b) DS maps c) Detected objects

Fig.9. Unsupervised object detection with our Deep Salience model. Red rectangles denote the object-like regions, and green ones highlight the most salient parts.

measurement. Hence, the dataset is well suited to our purpose of unsupervised object detection.

Our algorithm was implemented in MATLAB. We ran our code on the dataset, and compared the results against 11 state-of-the-art unsupervised visual salience models. The methods for comparison include: SR (Hou & Zhang 2007), IT (Itti *et al* in 1998), GB (Harel *et al* 2006), MZ (Ma & Zhang 2003), LC (Zhai & Shah 2006), FT (Achanta *et al* 2009), CA (Goferman *et al* 2010), HC and RC (Cheng *et al* 2011), MI (Jiang & Crookes 2012) and SF (Perazzi *et al* 2012). The following are our experimental results.

Visual Comparison

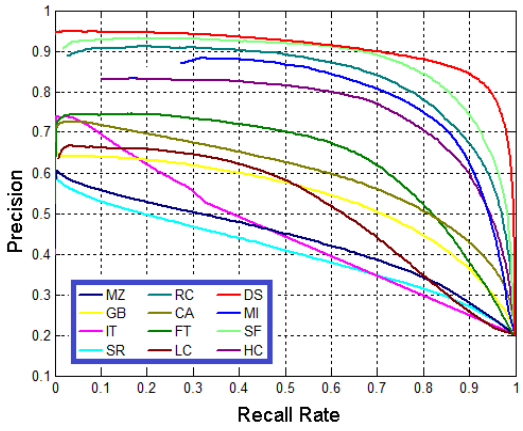
Fig.8 provides a visual comparison of the various methods using a number of sample images. Most of the state-of-the-art methods failed to find the salient objects in these test images. In comparison, we can see that our Deep Salience method using sMRF can robustly highlight salient objects from their background.

Fig.9 shows several more examples. By applying thresholds of 2*mean and 3*mean respectively to the salience maps, we have firstly identified object-like regions (shown as red rectangles) and then, even further, attempted to identify their salient parts (shown as green rectangles).

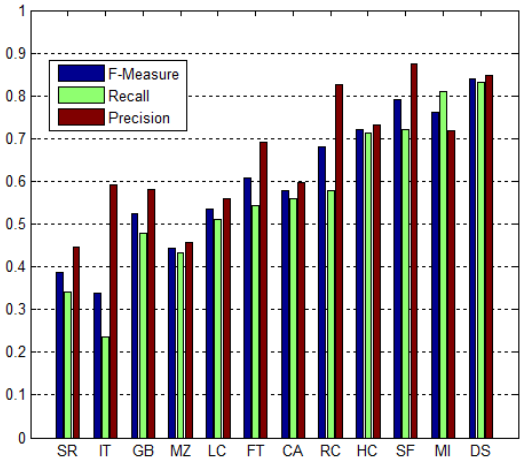
The success of our sQRF-based object detection can be attributed to the use of hierarchical analysis that is able to capture the global spatial structures of an image. However, we note that we did not build any contrast-based salience model explicitly. Instead, object detection is a natural outcome of maximizing the belief iteratively over the sQRF.

Statistic Evaluation

Fig.10-a shows the typical statistical results of precision-recall curves for the EPFL dataset. The curves are computed in the same way as reported by previous work. Here, precision corresponds to the percentage of salient pixels correctly assigned, while recall rate corresponds to the fraction of detected salient pixels in relation to the number of ground truth salient pixels. As shown in Fig.10-a, our method achieves the best overall precision and recall rates of all the compared methods, and consistently outperforms previous salience models in term of precision and recall rates. It is also observed that the DS curve goes up much steeper than other methods.



a) Precision-recall test



b) F-measure test

Fig.10. Evaluation on the EPFL database. a) Precision-recall test: DS model consistently achieved the best prevision/recall rates; b) F-measure test: DS attained the best scores in F-measure and recall, and the precision after thresholding was exceeded only by RC and SF.

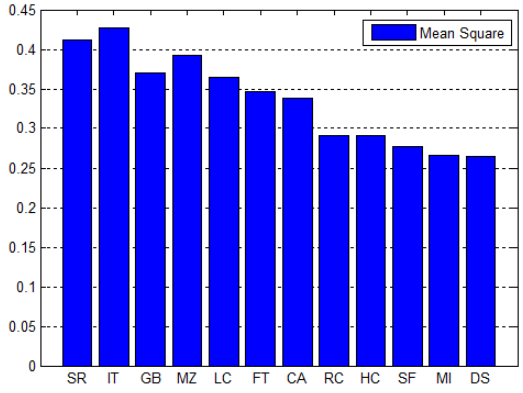


Fig.11. Mean square error (MSE) test.

Fig.10-b shows the results of *F*-measure, which combines both precision and recall to evaluate salience cut. Here, we use the adaptive threshold similar to that proposed by Achanta (2009), defined as twice the mean salience of the image. Appling this threshold to the salience map, we can get the precision and recall of its binary cut, and then the *F*-measure can be subsequently computed by:

$$F = \frac{\text{Precision} \times \text{Recall}}{(\text{Recall} + \text{Precision})} \qquad (6)$$

From Fig.10-b, we can see that our *Deep Salience* achieves the best *F*-measure score than all the other methods in the comparison.

Precision and recall measures do not consider true negative salience assignment, i.e., the number of pixel correctly marked as non-salient. Moreover, the quality of the weighted, continuous saliency maps may be of higher importance than the binary masks from psychological view of visual attention. To measure how successful a method is in the detection of non-salient background regions, we also carried out the comparison using mean square error (*MSE*) between the continuous saliency map *S* (before thresholding) and the binary ground truth *G*. *MSE* measure is then defined as,

$$MSE = \sqrt{\frac{1}{W \times H} \sum_x \sum_y \|S(x,y) - G(x,y)\|^2} \qquad (7)$$

where W and H are the width and the height of the respective saliency map and ground truth image.

Figure 11 shows that our method also outperforms other approaches in terms of the *MSE* measure, which provides a better estimate of the dissimilarity between the saliency map and ground truth. Results have been averaged over all images in the test dataset.

Conclusion

In conclusion, a new salience model has been successfully developed for unsupervised object detection via its deep belief propagation along the pyramid of successive Markov Random Fields (sMRF). The hierarchical structure of sMRF helps overcome the problem of background textures in belief propagation and also speeds up the convergence in multi-layer MRFs. It is also worth noting that our maximum belief propagation in QRF differs from typical mean-field approximation in that it employs a dynamic selection of fuzzy edges in its random fields, making it possible to guarantee that the belief propagation will not bleed over region borders in an image. Our benchmark experiments successfully validated that the proposed *Deep Salience* method consistently achieved better rates in precision-recall curves, and also attained better scores in both *F*-measure and *MSE* tests.

This work was initially inspired by the observed fact that visual salience arises during the propagation of visual stimuli along human visual cortex. Unlike most previous methods that are based on algorithmic math computation of various global contrasts, our salience estimation is a natural outcome of visual signal propagation over sMRF, which is more like the way how our human vision system processes the input visual signals.

References

Achanta, R.; Hemami, S.; Estrada, F.; Susstrunk, S. 2009. Frequency-tuned salient region detection. In *CVPR'09*, 1597–1604.

Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. 2011. Contour detection and hierarchical image segmentation. In *IEEE TPAMI*, 33(5): 898-916.

Birkhoff, G.; Neumann, J. V. 1936. The Logic of Quantum Mechanics. *Annals of Mathematics*, 37: 823–843.

Blue, M.; Bush, B. 2002. Unified approach to fuzzy graph problems. *Fuzzy Sets and Systems*, 125(3):355–368.

Cheng, M.; Zhang, G.; Mitra, N. J.; Huang, X.; Hu, S. 2011. Global Contrast based Salient Region Detection. In *CVPR'11*, 409-416.

Dalal, N.; Triggs, B. 2005. Histograms of oriented gradients for human detection. In *CVPR'05*.

Ekstrom, L. B.; Roelfsema, P. R.; Arsenault, J. T.; Bonmassar, G.; Vanduffel, W. 2008. Bottom-up dependent gating of frontal signals in early visual cortex. *Science*, 321(5887): 414-417.

Farabet, C.; Couprie, C.; Najman L.; LeCun, Y. 2013. Learning hierarchical features for scene labeling. In *IEEE TPAMI*, 35(8): 1915-1929.

Goferman, S.; Zelnik-Manor, L.; Tal, A. 2010. Context-aware saliency detection. In *CVPR'10*.

Harel, J.; Koch, C.; Perona, P. 2006. Graph-based visual saliency. In *NIPS*, 545–552.

Hou, X.; Zhang, L. 2007. Saliency detection: A spectral residual approach. In *CVPR'07*, 1–8.

Hinton, G. E.; Salakhutdinov, R. 2006. Reducing the dimensionality of data with neural networks. In *Science*, 313(5786): 504-507.

Itti, L.; Koch, C. 2001. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2:194–203.

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

Jegelka, S.; Bilmes, J. 2011. Submodularity beyond submodular energies: Coupling edges in graph cuts. In *CVPR'11*, 1897-1904.

Jiang, R.; Crookes, D. 2012. Visual saliency estimation through manifold learning. In *AAAI'12*.

Judd, T. 2009. Learning to predict where humans look. In *ICCV'09*.

Kohli, P.; Osokin, A.; Jegelka, S. 2013. A principled deep random field model for image segmentation. In *CVPR'13*.

Koch, C.; Ullman, S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227.

Kosko, B. 1986. Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, 24(1):65-75.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. In *IJCV*, 60(2): 91-110.

Ma, Y. F.; Zhang, H. J. 2003. Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia'03*, 374-381.

Perazzi, F.; Krähenbül, P.; Pritch, Y.; Hornung, A. 2012. Saliency Filters: Contrast Based Filtering for Salient Region Detection. In *CVPR'12*.

Peng, H.; Li, B.; Ji, R.; Hu, W. 2013. Salient object detection via low-rank and structured sparse matrix decomposition. In *AAAI-13*.

Rother, C.; Kolmogorov, V.; Blake, A. 2004. Grabcut–Interactive foreground extraction using iterated graph cuts. In *ACM Trans. Graph.*, 23(3): 309-314.

Salzenstein, F.; Collet, C. 2006. Fuzzy Markov Random Fields versus Chains for Multispectral Image Segmentation. *IEEE TPAMI*, 28(11): 1753 - 1767.

Sun, D.; Wulff, J.; Sudderth, E. B.; Pfister, H.; Black, M. J. 2013. A fully-connected layered model of foreground and background flow. In *CVPR'13*, 2451-2458.

Viola, P.; Jones, M. 2001. Robust real-time object detection. In *ICCV'01*.

Vedaldi, A.; Gulshan, V.; Varma, M.; Zisserman, A. 2009. Multiple kernels for object detection. In *ICCV'09*.

Yedidia, J. S.; Freeman, W. T.; Weiss, Y. 2005. Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms. *IEEE Trans. Information Theory*, 51:2282-2312.

Zhai, Y.; Shah, M. 2006. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*'06, 815-824.