Multi-Scale Saliency using Local Gradient and Global Colour Features

Christopher Cooley
Ulster University
Northland Rd
Derry/Londonderry, BT48 7JL
+44(0) 28 7167 5522
cooley-c3@ulster.ac.uk

Sonya Coleman
Ulster University
Northland Rd
Derry/Londonderry, BT48 7JL
+44(0) 28 7167 5030
sa.coleman@ulster.ac.uk

Bryan Gardiner
Ulster University
Northland Rd
Derry/Londonderry, BT48 7JL
+44(0) 28 7167 5081
b.gardiner@ulster.ac.uk

Bryan Scotney
Ulster University
Cromore Road
Coleraine, BT52 1SA
+44(0) 28 7012 4648
bw.scotney@ulster.ac.uk

ABSTRACT

In this paper, the issue of scale is addressed in the context of salient object detection. To date, many single scale models have been proposed for detecting salient objects within a scene. Scale is a fundamental problem within image processing, and therefore, multiple scale techniques are investigated and evaluated, as well the presentation of a novel multi-scale saliency model. The proposed model is compared with two state-of-the-art multi-scale saliency algorithms and qualitatively evaluated with respect to algorithmic accuracy and efficiency on the publicly available MSRA10K salient object dataset.

CCS Concepts

Computing Methodologies→ Artificial Intelligence→ Computer Vision→ Computer Vision Problems→ Object Detection

Keywords

Multi-scale Salient Object Detection; Hierarchical Saliency; Scale; Super-pixels; Salient Features.

1. INTRODUCTION

Visual saliency refers to the stimulus that causes an object/region to stand out, and therefore capture human attention. The Human Visual System (HVS) selectively processes visual data, before allocating attention to areas of interest, prior to further processing [1]. Scale is a fundamental problem within image processing. Within saliency detection the concept of scale has to be considered when extracting meaningful features from images. The strength of a feature depends on the scale at which it is detected. Therefore, some features become insignificant at particular scales. Issues have been identified in a number of saliency models, which struggle when dealing with scenes containing small-scale highly contrasting textures/patterns [2]. The posed question of this research is, what impact does scale have on salient object detection, and can any improvements be gained via different techniques? When calculating saliency at multiple scales, computation costs become a consideration, especially when global operations are implemented e.g. global colour. To address the aforementioned concerns, a number of scaling techniques are evaluated, and a novel multi-scale saliency algorithm is proposed. With respect to computational performance, super-pixels have been incorporated into the proposed model using the Simple Linear Iterative Clustering (SLIC) algorithm [3]. The focus of this research is low-level mathematical principles applied within saliency detection rather than shallow or deep learning. The remainder of the paper is structured as follows. Section 2 summarises relevant works. Section 3 outlines a number of scaling techniques, with the proposed multi-scale model presented in Section 4. Evaluation is undertaken in Section 5 and conclusions drawn in Section 6.

2. RELATED WORK

Many machine learning approaches have been applied to the challenge of detecting salient objects, for example [4] and [5]. However, the focus of this paper is bottom-up multi-scale saliency models, rather than deep or shallow learning models. Hierarchical Saliency Detection [2], introduces a novel approach to produce multiple layers of the input image. Each layer is the result of a merging process, meaning each layer includes different levels of detail from fine to coarse. Saliency cues are calculated on each layer, then fed into a hierarchical inference model to calculate the final saliency map. Visual Saliency Detection based on Gradient Contrast and Colour Complexity [6] adopts a Gaussian pyramid to create three image scales. At each scale, saliency features are calculated, then linearly combined. In the calculation of morphological gradient, three sizes of structuring elements are employed as a means of computing multi-scale gradient. In [7] a bottom-up multi-scale model is proposed using super-pixels as a scaling mechanism. Background and foreground priors are employed to calculate saliency. In [8], scale is utilised by enlarging the patch sizes at which the dissimilarity measure is calculated. A multi-scale super-pixel approach is presented in [9].

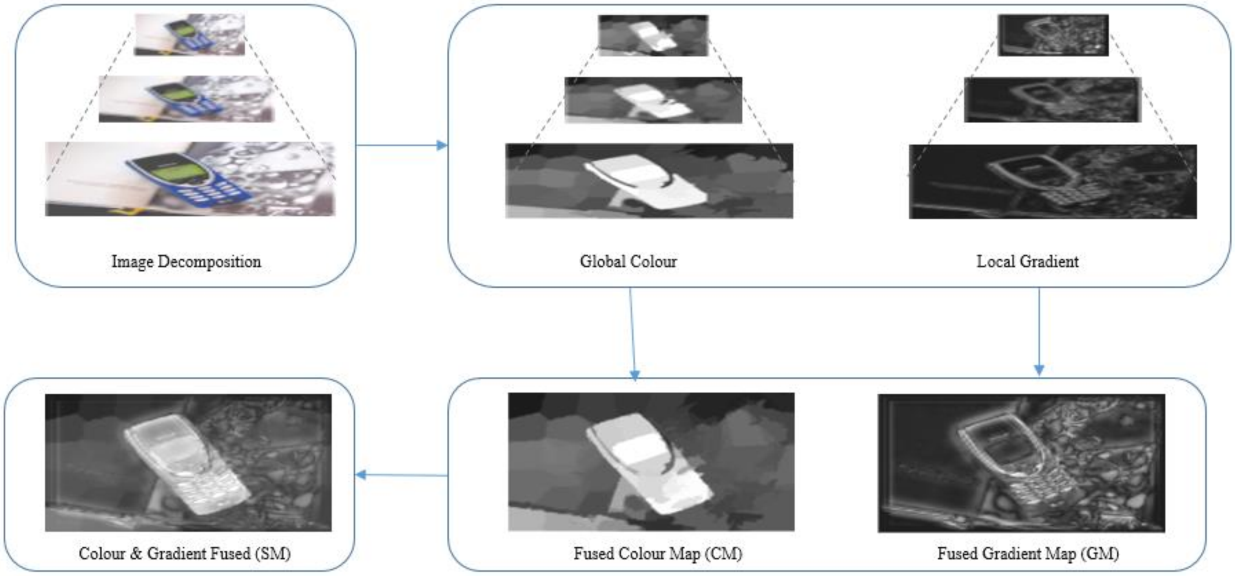

Figure. 1. Overview of multi-scale saliency model pipeline.

The calculation of features to determine saliency is often computationally expensive. As a result, many algorithms have adopted clustering/segmentation, to group pixels and form perceptually meaningful regions, before performing feature extraction. A watershed-like method was implemented in [2], whereas super-pixels were adopted in [6]. As a means of improving computation time, the presented model will incorporate the use of super-pixel regions. A number of different methods have considered scale as a means of improving salient object detection. Within this work, gradient is employed as a feature cue, in conjunction with evaluating different multi-scale approaches similar to those used in [2, 6, 8 and 9].

3. SCALING TECHNIQUES

The issue of scale has been approached with a number of different techniques, a few of which are outlined in this section. The well-known pyramid scheme is used to represent images hierarchically. Input image I of size $M \times N$, follows a repeated process of smoothing and resampling.

Super-pixels encode important information regarding object shape and boundary, aiding segmentation. However, the challenge is choosing the optimal number of super-pixels, which has a direct impact on the success of saliency features. Figure 2 shows an image partitioned with a varying number of super-pixels. At each scale of super-pixel, different information regarding the salient object is captured. As in [7, 9, 10], image I is segmented using multiple levels of super-pixels K. To calculate multi-scale gradient, multiple scales of the Near-Circular operator [11] are used namely operators of size $3 \times 3,\ 5 \times 5$ and 7×7, referred to as multi-scale mask and super-pixel approach (MSM&SP)

Methods [2, 6, 9] as detailed in Section 2 are also evaluated as scaling techniques in Section 5. The work in [2] proposes a layering hierarchical scheme, with each image layer containing different levels of detail. At each layer a mean filter is used to calculate a scale value, with values less than a threshold being combined with neighbour regions. The algorithm in [6] processes three scales of images, produced by a Gaussian Pyramid. Colour complexity and morphological gradient are calculated at each scale, before producing the final saliency map. In [9] multi-scale segmentations of super-pixels are computed. Various Gaussian smoothing variables are utilised in the generation of both coarse and fine results.

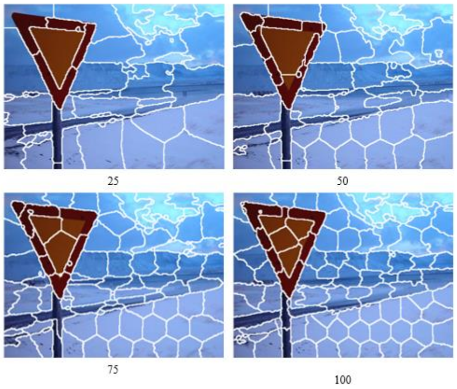

Figure. 2. Image with varying number of super-pixels.

4. MULTI-SCALE SALIENCY APPROACH

Within this section a novel multi-scale salient object detection model is presented entailing four main steps. First, the input image is decomposed into three different scales $[R, R/2,\ R/4]$, using a pyramid approach, where R is the original scale of the input image. Within this work, image I is subsampled into three scales by $M/2$ and $N/2$. As the gradient operator employed in this research for feature detection has built-in Gaussian smoothing, it is therefore not included as part of the image decomposition, resulting in computational speed-up. Next, each image layer is over-segmented with super-pixels. Feature cues are then computed on each layer, and finally amalgamated to obtain the resultant saliency map. The model pipeline is shown in Figure 1.

4.1 Super-Pixel Segmentation

After the original image is decomposed into three different levels, each image level is segmented into super-pixels using the Simple Linear Iterative Clustering (SLIC) algorithm [3]. SLIC clusters pixels with similar colour values in close proximity, permitting images to be processed in a region-based manner, thus improving efficiency. The challenge with processing regions is achieving the correct balance between computation and accuracy. To choose an optimal number of super-pixels, colour contrast was computed across 100 randomly selected images, determining the mean accuracy and runtime while varying the number of super-pixels generated throughout the image domain. Results are presented in Table 1.

Table 1. Mean accuracy and runtime by varying number of super-pixels across 100 images.

Super-Pixels	*Accuracy*	*Runtime (secs)*
15	91.18	0.81
20	91.18	0.82
30	91.49	0.92
40	92.03	0.98
45	**92.82**	**0.92**
50	92.81	0.92
60	92.59	1.01
70	92.59	1.01
80	92.46	1.15
90	92.41	1.23
100	92.40	1.15

The number of super-pixels was varied from $15 - 100$. The experiment found that accuracy, as well as runtime, generally increased with a higher number of super-pixels. However, after 45 super-pixels accuracy started to decline while runtime continued to rise. Within the SLIC algorithm, the number of super-pixels (K) is manually selected. On all three image scales, K was chosen as 45, as this scored the highest accuracy. It should be noted, although K is manually selected, the actual number of super-pixels may vary slightly depending on the image, observed during experimentation.

4.2 Feature Cues

Low-level feature cues are an essential part of saliency detection. Algorithms normally adopt two or more features, as one typically doesn't suffice when processing vast numbers of colours, objects, backgrounds and lighting. This section details the feature cues derived for the detection of salient objects.

4.2.1 Global Colour

One of the main visual features that fascinates human attention is colour. Pixels/regions that have a high contrast to their surroundings are considered to be salient [12]. Many models compare pixels/regions to their neighbours which is known as local contrast. Feature maps from local colour contrast tend to be noisy and mainly highlight the edges of the salient object, whereas global contrast computes the contrast of a pixel/region in relation to all of the remaining pixels/regions within an image. Global colour contrast (CC) of a super-pixel i at each scale is defined as:

$$CC_{(i)} = \sum_{j=1}^{N} \tau(c_{(i)}, c_{(j)}) \sqrt{(D_{(i)} - D_{(j)})^2} \qquad (1)$$

where N is the total number of super-pixels and $D_{(i)}$ and $D_{(j)}$ are the average $L^*\ a^*\ b^*$ values of super-pixels i and j respectively. $\tau(c_{(i)}, c_{(j)})$ is a weighting term controlling the range of colour contrast feature, calculated as:

$$\tau(c_{(i)}, c_{(j)}) = \exp\left(-\frac{1}{0.125}\left\|c_{(i)} - c_{(j)}\right\|^2\right) \qquad (2)$$

where $c_{(i)}$ and $c_{(j)}$ are the centre positions of super-pixels i and j respectively.

4.2.2 Local Gradient

Having evaluated a family of Gaussian based derivate operators, specifically the Linear Gaussian [13], Bilinear Gaussian [14] and the Near-Circular [11], in the context of saliency. The Near-Circular operator [11] was found to be best suited for saliency detection, outperforming the other compared operators. Therefore, the 7×7 Near-Circular operator is adopted, and gradient contrast calculated across a local neighbouring region of $[9 \times 9]$ pixels. Gradient contrast (GC) per scale level is formulated as:

$$GC_{(i)} = \sum_{i=1}^{N}\sum_{j=1}^{N}\left\|g_{(i,j)} - g_{(n)}\right\| \qquad (3)$$

where N is the total number of pixels within the neighbourhood region and $g_{(i,j)}$ and $g_{(n)}$ are gradient magnitude value of pixel i, j and the sum of gradient values across a neighbourhood n respectively.

4.3 Scale and Feature Fusion

Fusion of features is a key step in any saliency algorithm to obtain a final saliency image/map. With the proposed model implemented at multiple scales, CC and GC require amalgamation at each scale. Gradient and colour features are fused independently with their respective multi-scale maps, likened to an inverse pyramid scheme. Prior to this, feature maps are re-scaled to the original size and fused by algorithmically summing each feature map output, such that the fused colour scale map (CM) can be defined by:

$$CM_{(x,y)} = \sum_{r=1}^{3} CC_{(x,y)}^{r} \qquad (4)$$

where $CC_{(x,y)}^{r}$ is the colour contrast value of the pixel at coordinate (x, y) at each scale r. After the scaled feature maps are merged by summation of corresponding pixels at each scale, the resultant fused feature cues need to be integrated to form the final saliency map. The algorithm's final saliency map SM can be calculated as:

$$SM = (\alpha * CM) + (\beta * GM) \qquad (5)$$

where α is set to 0.7 and β is 0.6. It was empirically determined this combination of weights yielded the best results. CM and GM (as depicted in Figure 1) are the colour and gradient fused scaled feature maps respectively.

5. EVALUATION

To evaluate the proposed saliency approach, different metrics were used as outlined in Section 5.1. Firstly, the proposed approach was implemented on a single-scale referred to as Single Scale Saliency (SSS), progressing to multi-scale, comparing with a number of techniques. These techniques included, using a pyramid scheme to decompose the image into different scales, and implementing the proposed feature cues but scaling the number of super-pixels rather than the image [7, 9]. In the latter approach, the multi-scale gradient feature entailed using multiple scales of near-circular operator masks namely $\mathbf{3 \times 3}$, $\mathbf{5 \times 5}$, $\mathbf{7 \times 7}$ (MSM&SP). The final proposed model is evaluated against two state-of-the-art multi-scale saliency approaches: Visual Saliency Detection based on Gradient Contrast and Colour Complexity [6] (VSD) and Hierarchical Saliency Detection [2] (HSD), on the publicly available MSRA10K salient object dataset [14].

5.1 Evaluation Metrics

Before calculating evaluation metrics, the resultant saliency map SM is binarised by varying a threshold from $[0, 255]$. At each threshold, the Accuracy measure A is calculated, which is the percentage difference when comparing each pixel of the binarised saliency map BM with the associated ground truth mask M. This determines how successful a model is at correctly labelling pixels as salient or non-salient and is outlined as:

$$A = \frac{(tp + tn)}{(tp + tn + fp + fn)} \tag{6}$$

where tp, tn, fp and fn are defined as true positives, true negatives, false positives and false negatives respectively.

Precision and recall are used to evaluate saliency models due to their similarity with a linear classification problem. The precision value refers to the ratio of correctly assigned salient pixels against all extracted regions. Recall measures the percentage of truly salient regions the algorithm was able to correctly label.

Receiver operating characteristics (ROC) curve reports the false positive rate (FPR) against the true positive rate (TPR) at different thresholds and each is calculated by:

$$TPR = \frac{BM \cap M}{BM}, \ FPR = \frac{BM \cap M}{M} \tag{7}$$

From the ROC curve, the area under ROC curve (AUROC) can be calculated, with a perfect model scoring an AUROC of 1 and an AUROC score of 0.5 is equated to guessing.

5.2 Results

The proposed model was tested on a single scale model as well as different multi-scale techniques, as well as evaluated against two state-of-the-art multi-scale saliency approaches, namely, HSD [2] and VSD [6]. As seen in Table 2, the maximum and mean of each algorithm's accuracy are recorded. Hierarchical saliency detection scored the highest in both, with the proposed pyramid multi-scale approach following closely. The highest AUROC score was recorded by the proposed model scoring 0.9321, with visual saliency detection coming in second with a 0.8967 score. Runtime was recorded in seconds measuring algorithm efficiency. The single scale saliency (SSS) approach recorded the fastest average runtime, which is expected as every other approach is processing multiple scales. The proposed model closely followed, recording a runtime of 0.4 seconds on average. The precision/recall and ROC curves are represented in Figure 3. Figure 4 shows a visual comparison of each algorithm, where our approach can be seen to detect and highlight the internal regions of the salient object, as well as preserving the edges with fine scale details. In particular, images on rows one, three and four, show our method outperforming the other techniques in terms of highlighting the entire salient object consistently. Some techniques can be seen to completely miss certain internal small-scale regions of the salient object, specifically, VSD, MSM&SP and SSS.

6. CONCLUSION

This paper investigated gradient information as a feature for use in salient object detection. The Near-Circular derivative operator was utilised for the calculation of gradient feature, within the proposed model. The presented algorithm combines local gradient contrast with global colour contrast. The algorithm was implemented on a single-scale, as well as on two different multi-scale approaches. A study was also completed for choosing the optimal number of super-pixels, found to be 45.

The proposed model is evaluated against two state-of-the-art hierarchical algorithms, outperforming them in terms of ROC, AUC and runtime. Further investigation is required to improve the computational efficiency of the proposed algorithm for real-time usage, as seen in the presented results. Other feature cues such as depth, texture and motion will be considered as means of improving the robustness of the proposed approach, with a view to extending the algorithm for use in videos.

7. REFERENCES

[1] A. Borji, M. M. Cheng, H. Jiang, and J. Li, "Salient Object Detection : A Benchmark," IEEE Trans. Image Process., vol. 24, no. 12, pp. 5706–5723, 2015.

[2] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical Saliency Detection," in Computer Vision and Pattern Recognition (CVPR), 2013.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. S.¨sstrunk, "SLIC Super-pixels Comparedto State-of-the-Art Superpixel Methods," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274–2281, 2012.

[4] G. Li and Y. Yu, "Deep Contrast Learning for Salient Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 478–487, 2016.

[5] X. Wang, H. Ma, X. Chen, and S. You, "Edge Preserving and Multi-Scale Contextual Neural Network for Salient Object Detection," IEEE Trans. Image Process., vol. 27, no. 1, pp. 121–134, 2018.

[6] W. Li, J. Qui, and X. Li, "Visual Saliency Detection based on Gradient Contrast and Color Complexity," in International Conference on Internet Multimedia Computing and Service, 2015, pp. 1–5.

[7] X. Lin, Z. Wang, L. Ma, and X. Wu, "Saliency Detection via Multi-Scale Global Cues," IEEE Trans. Multimed., vol. 21, no. 7, pp. 1646–1659, 2019.

[8] S. Anwar, Q. Zhaot, and M. F. Manzoor, "Saliency Detection using Parallel Non-Linear Integration of Color and Gradient using Covariances," in International Conference on the Innovative Computing Technology, 2014, pp. 197–201.

[9] N. Tong, H. Lu, L. Zhang, and X. Ruan, "Saliency Detection with Multi-Scale Super-pixels," IEEE Signal Process. Lett., vol. 21, no. 9, pp. 1035–1039, 2014.

[10] N. Mu, X. Xu, Y. Wang, and X. Zhang, "A Multiscale Superpixel-Level Salient Object Detection Model Using Local-Global Contrast Cue," J. Shanghai Jiaotong Univ., vol. 22, no. 1, pp. 121–128, 2017.

[11] B. W. Scotney, S. A. Coleman, and M. G. Hewon, "A Systematic Design Procedure for Scalable Near-Circular Gaussian Operators" in IEEE International Conference on Image Processing (ICIP 2001), pp. 844–847, 2001.

[12] Q. Zhang, J. Lin, Y. Tao, W. Li, and Y. Shi, “Salient object detection via color and texture cues,” Neurocomputing, vol. 243, pp. 35–48, 2017.

[13] S. A. Coleman and B. W. Scotney, “Image Feature Detection on Content-Based Meshes,” in IEEE International Conference on Image Processing (ICIP 2002), pp. 844–847, 2002.

[14] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu, “Global Contrast Based Salient Region Detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 3, pp. 569–582, 2015.

[15] B. W. Scotney, S. A. Coleman, M. G. Herron, and S. Engineering, “Device Space Design for Efficient Scale-Space Edge Detection,” in International Conference on Computational Science (ICCS 2002), pp. 1077–1086, 2002.

Table 2. Comparison of multi-scale saliency technique results

Approach	Max Acc.	Mean Acc.	AUROC	Runtime
Single-Scale Approach (SSS)	99.8%	91.6%	0.7108	0.3 secs
Proposed Pyramid Multi-Scale Approach (Ours)	99.8%	92.9%	0.9321	0.4 secs.
Multi-Scale Super-Pixel Approach (MSM&SP)	99.8%	92.4%	0.8673	0.7 secs.
Visual Saliency Detection (VSD)	99.7%	92.8%	0.8967	0.9 secs
Hierarchical Saliency Detection (HSD)	99.9%	96.3%	0.5866	0.6 secs

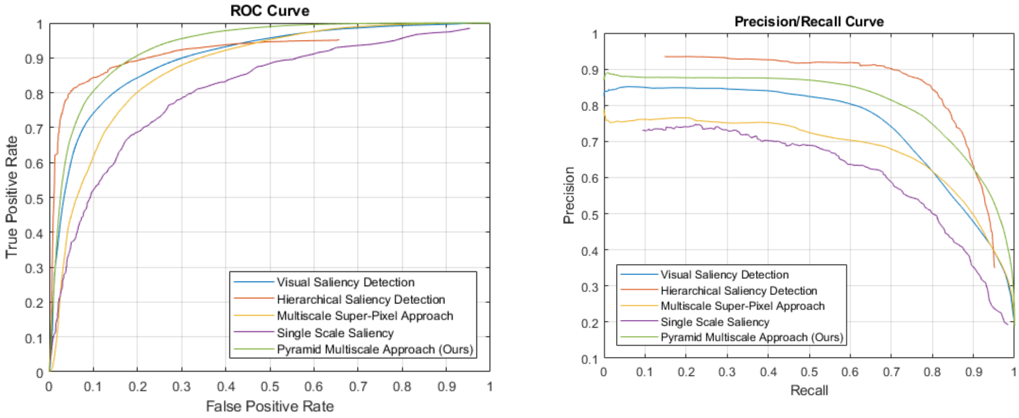

Figure. 3. ROC and Precision/Recall Curves

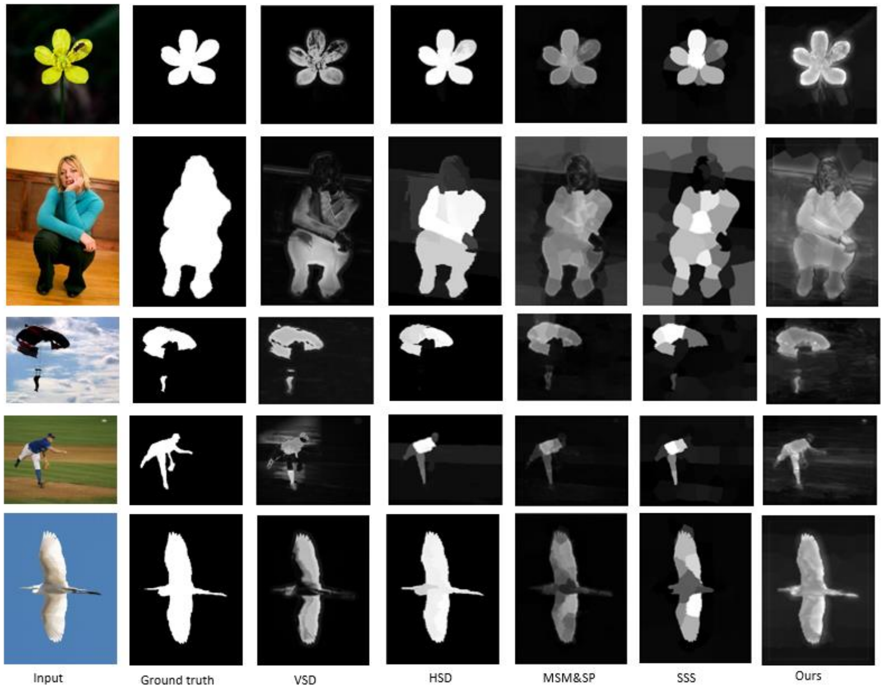

Figure. 4. Visual comparison of saliency approaches.