GM-W9: Two New Methods from ICCV 2013

Hongyang Li (李弘扬)

IIAU Lab, Dalian University of Technology

hongyangli2020@gmail.com

Supervisor: Prof. Huchuan Lu (卢湖川 教授)

November 1, 2013

ICCV 2013: A Statistical Snapshot. As of 29 Aug, 2013

Papers submitted[1]: 1629
Withdrawals and administrative rejections: 128
Accepted as Orals: 41 (2.52% oral acceptance rate)
Accepted as Posters: 413 (27.87% total acceptance rate)

Acceptance per Primary Subject Area:

Primary Subject Area	Submitted	Accepted	Acceptance rate
Low-level vision and image processing	132	28	21%

[1] http://www.cs.toronto.edu/ kyros/local_outgoing/ICCV-Final-Results/

ICCV 2013: A Statistical Snapshot. As of 29 Aug, 2013

Papers submitted[1]: 1629
Withdrawals and administrative rejections: 128
Accepted as Orals: 41 (2.52% oral acceptance rate)
Accepted as Posters: 413 (27.87% total acceptance rate)

Acceptance per Primary Subject Area:

Primary Subject Area	Submitted	Accepted	Acceptance rate
Low-level vision and image processing	132	28	21%

Saliency detection, about 12 Papers.

1. Efficient Salient Region Detection with Soft Image Abstraction
Mingming Cheng. Project Link

2. Benchmarking Computational Model of Visual Saliency
Ali Borji, Laurent Itti. *Oral.*

3. Category-Independent Object-level Saliency Detection
Yangqing Jia (UC Berkeley), Mei Han (Google Research).

[1] http://www.cs.toronto.edu/ kyros/local_outgoing/ICCV-Final-Results/

Introduction

From IIAU Lab, DLUT.

- Xiaohui Li, *et al.*

  Saliency Detection via Dense and Sparse Reconstruction

  ICCV 2013.

Segmenting Salient Objects from Images and Videos

Esa Rahtu[1], Juho Kannala[1], Mikko Salo[2], and Janne Heikkilä[1]

[1]Machine Vision Group, University of Oulu, Finland
[2]Department of Mathematics and Statistics, University of Helsinki, Finland

Abstract. In this paper we introduce a new salient object segmentation method, which is based on combining a saliency measure with a conditional random field (CRF) model. The proposed saliency measure is formulated using a statistical framework and local feature contrast in illumination, color, and motion information. The resulting saliency map is then used in a CRF model to define an energy minimization based segmentation approach, which aims to recover well-defined salient objects. The method is efficiently implemented by using the integral histogram approach and graph cut solvers. Compared to previous approaches the introduced method is among the few which are applicable to both still images and videos including motion cues. The experiments show that our approach outperforms the current state-of-the-art methods in both qualitative and quantitative terms.

Keywords: Saliency measure, background subtraction, segmentation

1 Introduction

Biological vision systems are remarkably effective in finding relevant targets from a scene [1]. Identifying these prominent, or salient, areas in the visual field enables one to allocate the limited perceptual resources in an efficient way. Compared to biological systems, computer vision methods are far behind in the ability of saliency detection. However, reliable saliency detection methods would be useful in many applications like adaptive compression and scaling [2, 3], unsupervised image segmentation [4, 5], and object recognition [6, 7].

Perhaps the most common approach to reduce scene clutter is to detect moving objects against a static background [8–10]. These methods have been very successful in many applications, but they have severe limitations in the case of dynamic scenes or moving cameras. These circumstances have been addressed by introducing adaptive background models and methods to eliminate camera movements [11, 12], but both of these are difficult problems and technically demanding. Moreover, the methods in this class are applicable only to video sequences, but not to still images where motion cues are not available.

A different approach is provided by supervised object detection techniques, which are aimed at finding particular categories like persons, tables, cars, etc. [13–15]. These methods have resulted in high performance, but the limitation is

Idea I

Given an image, we have feature matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$ and background templates $\mathbf{B} \in \mathbb{R}^{D \times M}$.

Incentive 1: PCA

- compute the bases of the template $\mathbf{B}$.
- dimensionality reduction: $\mathbf{P} \in \mathbb{R}^{D' \times N}$ (here $D' = 3$).

$$\mathbf{p}_i = \mathbf{U}_\mathbf{B}^T (\mathbf{x}_i - \bar{\mathbf{x}}) \tag{1}$$

$$\varepsilon_i^d = \|\mathbf{x}_i - (\mathbf{U}_\mathbf{B} \mathbf{p}_i + \bar{\mathbf{x}})\|_2^2 \tag{2}$$

where ε_i^d denotes how much different the i-th segment is from the background template.

Idea II

Incentive 2: Sparse Representation

- choose template $\mathbf{B}$ as the dictionary.
- find the sparse representation $\mathbf{Q} \in \mathbb{R}^{M \times N}$ via optimisation.
- use $\mathbf{Q}$ to reconstruct $\mathbf{X} = \mathbf{BQ}$.

$$\mathbf{q}_i = arg \min_{\mathbf{q}_i} \left(\frac{1}{2} \|\mathbf{x}_i - \mathbf{B}\mathbf{q}_i\|_2^2 + \lambda \|\mathbf{q}_i\|_1 \right) \tag{3}$$

$$\varepsilon_i^s = \|\mathbf{x}_i - \mathbf{B}\mathbf{q}_i\|_2^2 \tag{4}$$

where ε_i^s *also* denotes how much different the i-th segment is from the background template.

PCA vs. Sparse

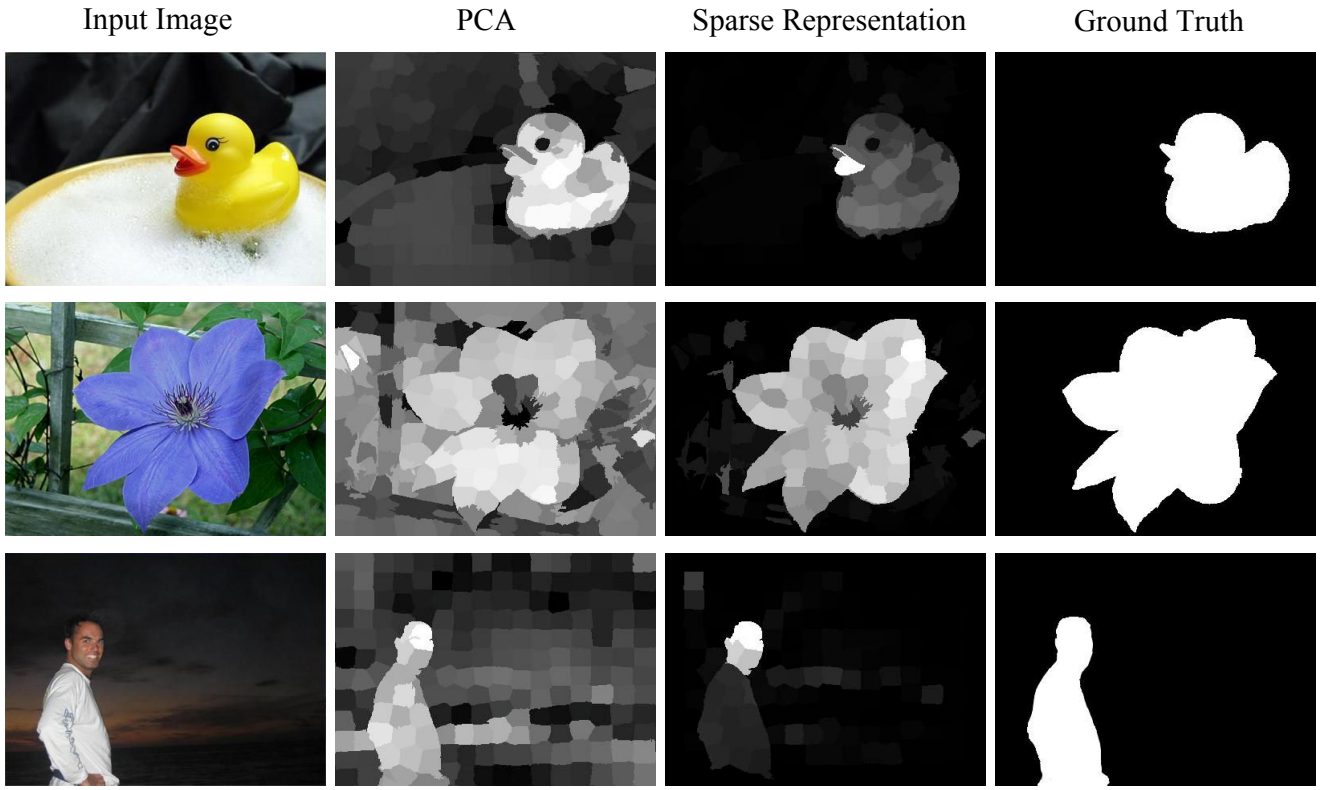


Figure: Saliency Map Comparison. (X. Li, *et al.* ©IEEE 2013)

PCA vs. Sparse

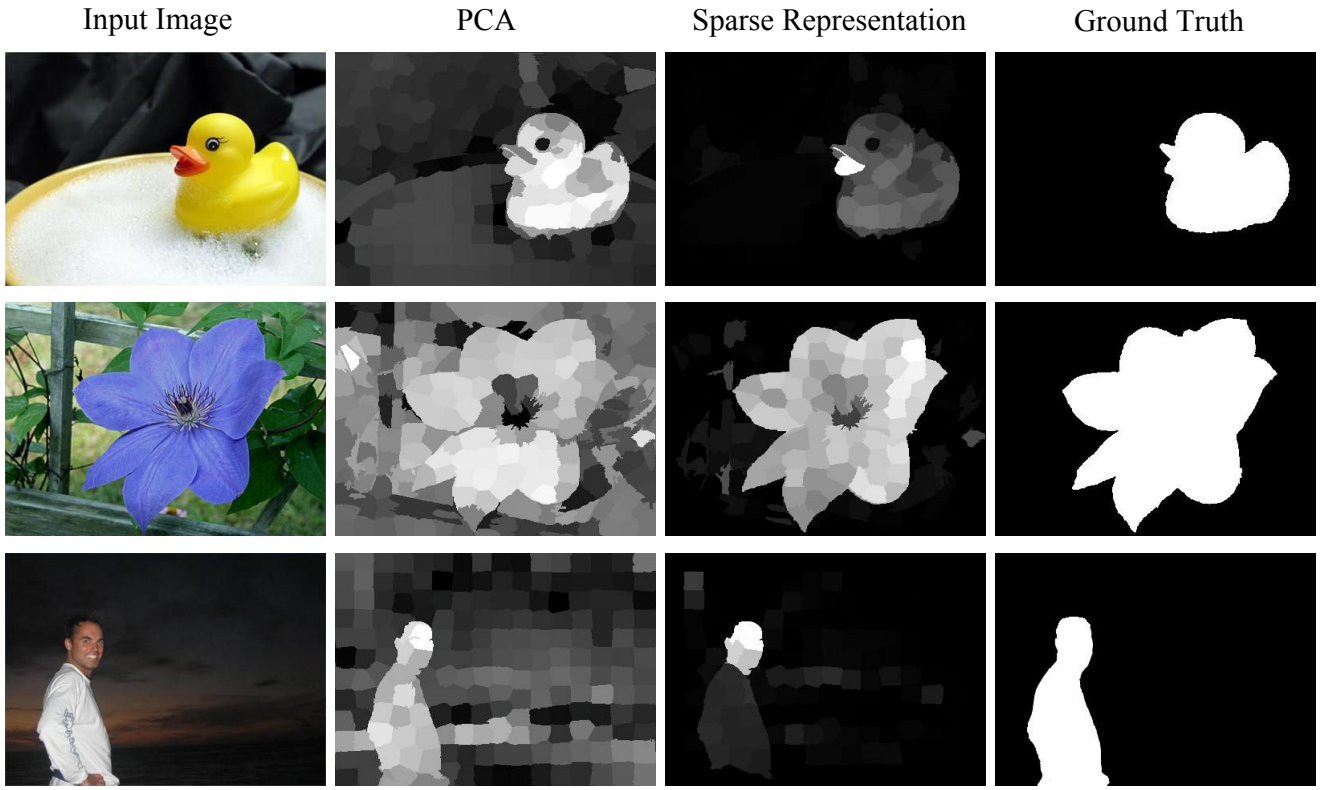


Figure: Saliency Map Comparison. (X. Li, *et al.* ©IEEE 2013)

As the author puts it:

1. PCA is robust to boundary salient objects; whereas
2. Sparse is more accurate to complicated background.

Refinement along the road

1. Saliency (recons. error) Update via Propagation

Apply k-means to cluster N segments(superpixels) into K clusters.

$$\widetilde{\varepsilon}_i = \tau \sum_{j=1}^{N_k} w_{ij} \widetilde{\varepsilon}_j + (1-\tau)\varepsilon_i \tag{5}$$

where N_k denotes the number of segments in the k-th cluster.

- Aim: to smooth saliency values and highlight the salient objects;
- Mechanism: sorted in descending order and processed sequentially;
- Other methods? like PageRank.

Refinement along the road

2. Pixel-level Saliency: Multi-scale and Object-biased

Let $E(z)$ indicate the saliency value of each pixel under different SLIC Alg. scales. Also we integrate an object-biased Gaussian model.

$$S(z) = E(z) \cdot G_o(z) \tag{6}$$

$$G_o(z) = \exp\left[-\left(\frac{(x_z - x_o)^2}{2\sigma_x^2} + \frac{(y_z - y_o)^2}{2\sigma_y^2}\right)\right] \tag{7}$$

Bayesian Integration of Saliency Maps

- **How to integrate two methods?**
- Incentive: Rahtu, ECCV10. Recall that

$$S(x) = P(H_0|F(x)) = \frac{P(F(x)|H_0)P(H_0)}{P(F(x)|H_0)P(H_0) + P(F(x)|H_1)P(H_1)} \quad (8)$$

$$\triangleq \frac{h_K(x)p_0}{h_K(x)p_0 + h_B(x)(1-p_0)} \quad (9)$$

Given two saliency maps S_i and S_j from either dense or sparse reconstruction, we treat S_i as the **prior** and the other S_j to compute the **likelihood**. Therefore, we obtain

$$P(F_i|S_j(z)) = \frac{P(S_j(z)|F_i)S_i(z)}{P(S_j(z)|F_i)S_i(z) + P(S_j(z)|B_i)(1-S_i(z))} \quad (10)$$

Bayesian Integration of Saliency Maps

Once we have $P(F_i|S_j(z))$, the final saliency measure is formulated as

$$S_B(z) = P(F_1|S_2(z)) + P(F_2|S_1(z)) \tag{11}$$

where the subscript 'B' stands for Bayesian integration.

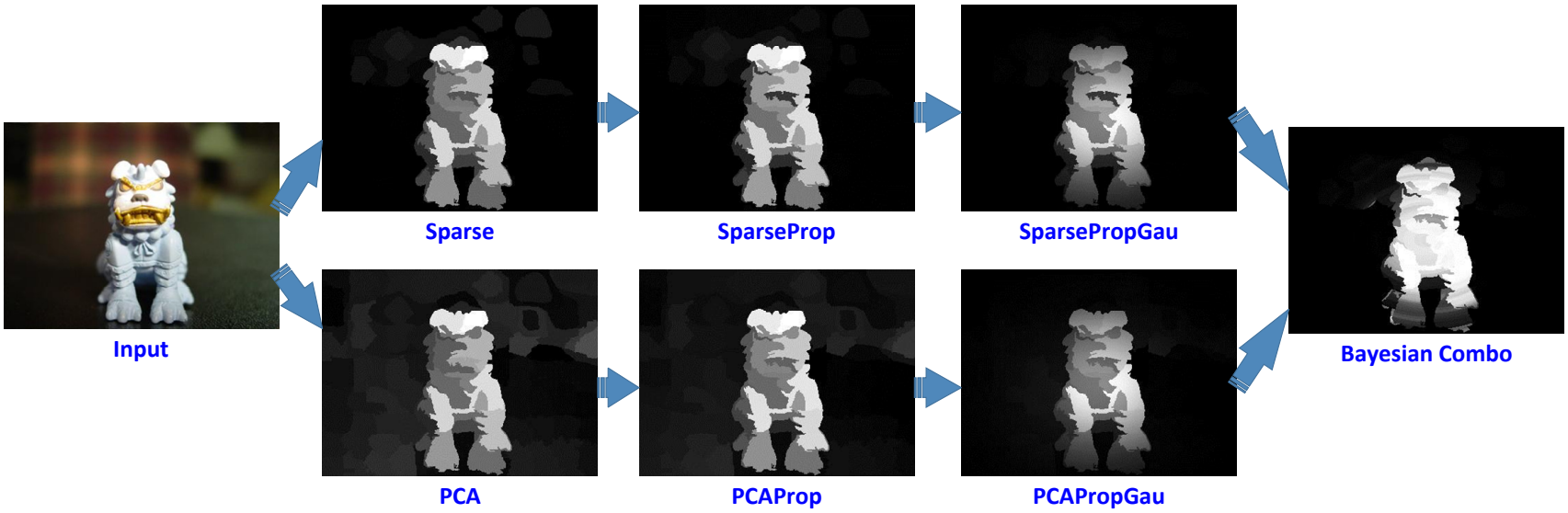


Figure: Saliency Maps on the whole.

Horizonal Comparison

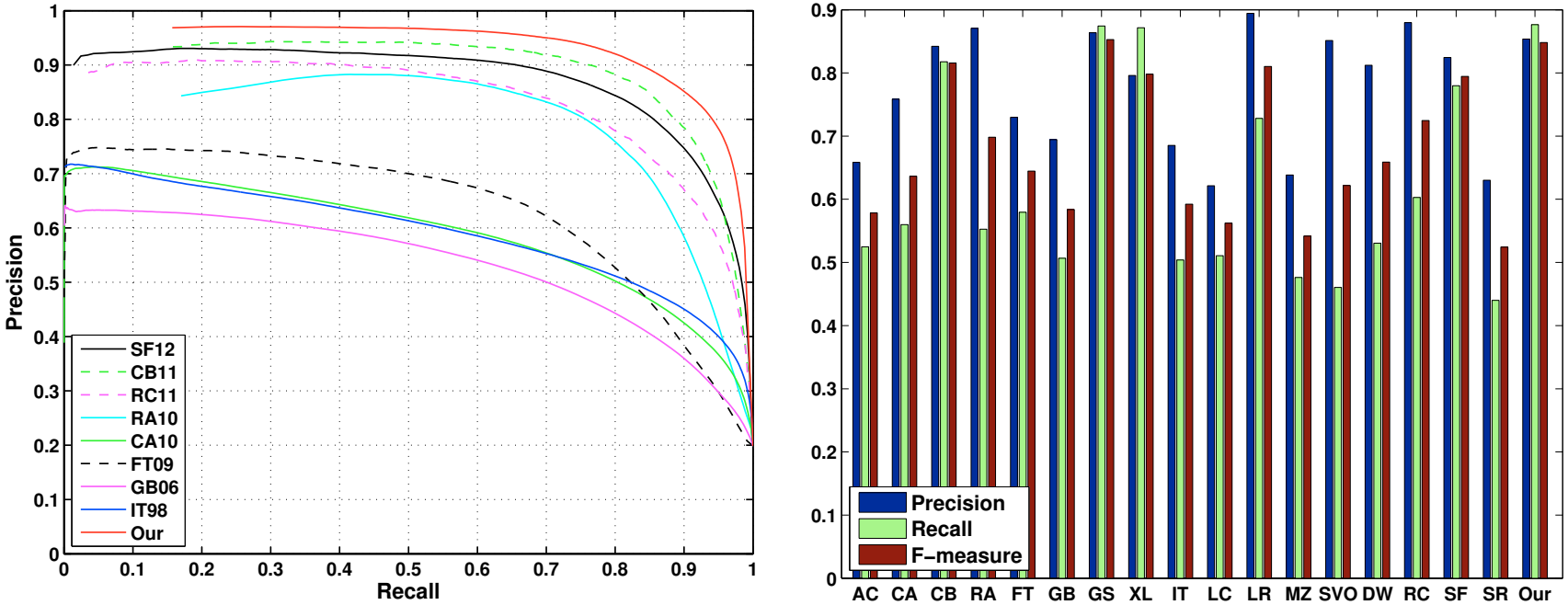


Figure: Performance of the proposed alg. compared with state-of-the-art methods on the ASD database. (X. Li, *et al.* ©IEEE 2013)

Introduction

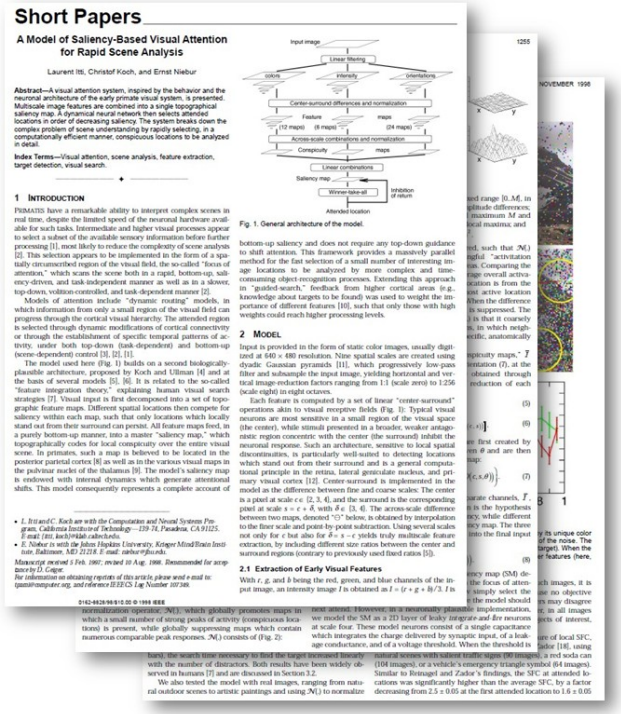
From Boston University.

J. Zhang and S. Sclaroff.

Saliency Detection: A Boolean Map Approach

ICCV 2013.

Project website:
http://cs-people.bu.edu/jmzhang/BMS/BMS.htm



Idea

▷ **Incentive:** A Boolean Map Theory of Visual Attention[2].
'An observer's momentary conscious awareness of a scene can be represented by a boolean map.'

[2] http://www.pashler.com/Articles/Huang_Pashler_PR2007.pdf

Idea

▷ **Incentive:** A Boolean Map Theory of Visual Attention[2].
'An observer's momentary conscious awareness of a scene can be represented by a boolean map.'

▷ **A(B) Computation:** A Gestalt principle for figure-ground segregation.
'Surrounded regions are more likely to be perceived as figures.'
Computationally, we employ **Flood Fill algorithm** to assign 1 within the surrounded regions and 0 to the rest of the map.

[2] http://www.pashler.com/Articles/Huang_Pashler_PR2007.pdf

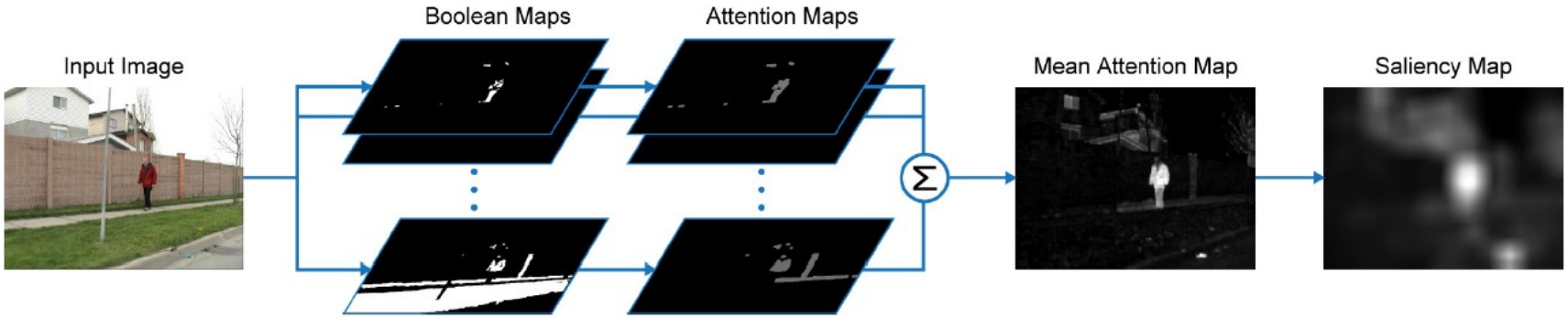


Figure: The Pipeline of Boolean Map Saliency

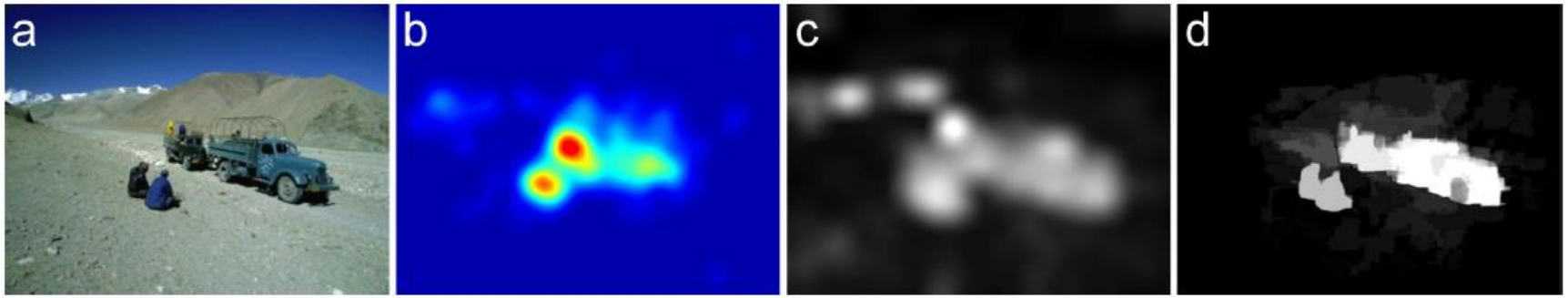
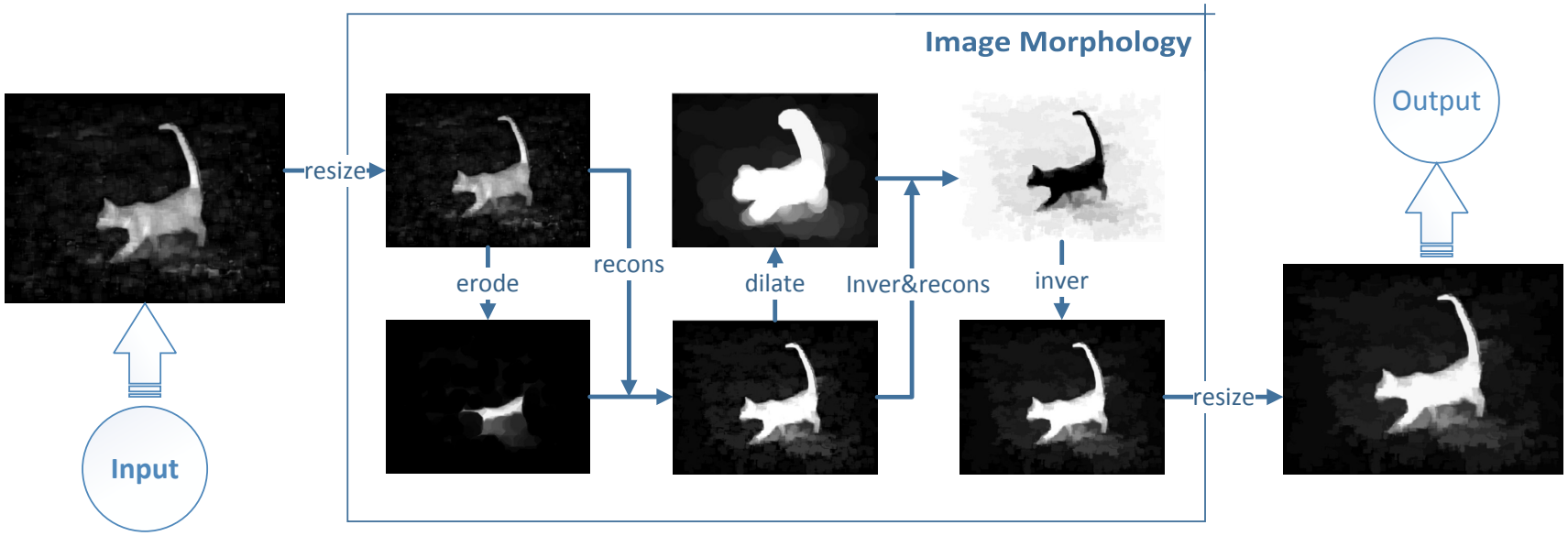


Figure: Saliency Maps

Algorithm 1 Boolean Map Saliency, $S = \texttt{BMS}(I)$

```
B = {}                                   ▷ B = {B_1, B_2, ··· , B_m, ··· , B_N}
for each color channel map {φ_k(I), k = 1, 2, 3} do
    for θ = 0 : δ : 255 do
        B = THRESH(φ_k(I), θ), B̃ = INVERT(B)
        add OPENING(B, ω_0) and OPENING(B̃, ω_0) to B
    end for
end for
for each B_m ∈ B do
    A_m = ZEROS(size(B_m))
    set A_m(i, j) = 1 if B_m(i, j) belongs to a surrounded region
    A_m = DILATION(A_m, ω_d1), A_m = NORMALISE(A_m)
end for
Ā = (1/N) Σ_{m=1}^{N} A_m
S = POST_PROCESS(Ā)                      ▷ for Eye Fixation, Salient Detection
return S
```

Post-processing for Salient Object Detection



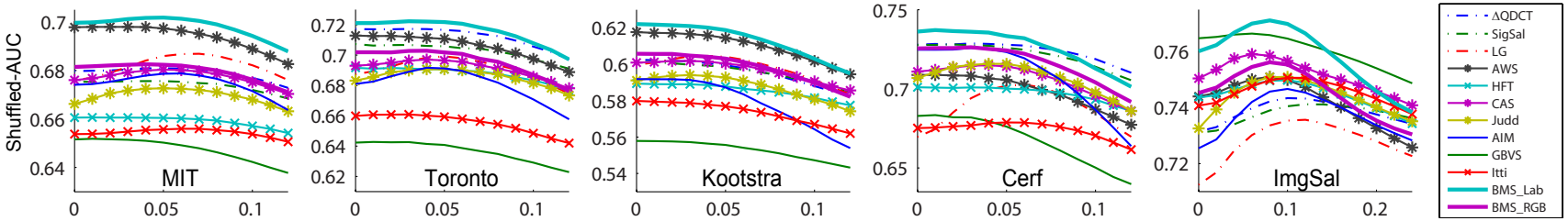


Figure 4: **Average Shuffled-AUC against the STD of Gaussian Blur**. X-axis represents the Gaussian blur standard deviation (STD) in image width and Y-axis represents the average shuffled-AUC score on one dataset.

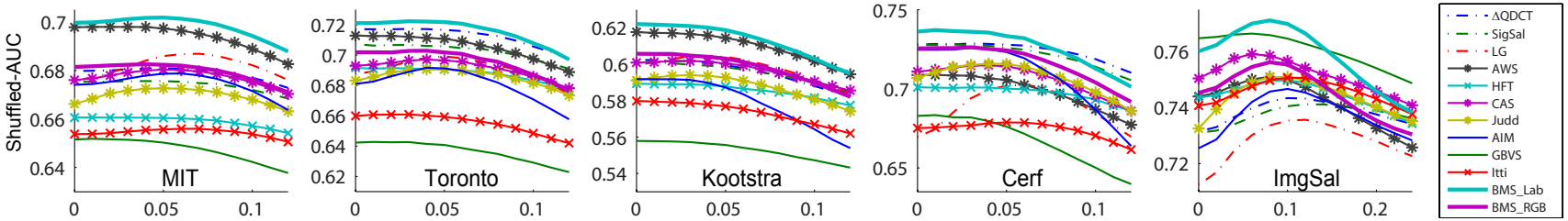
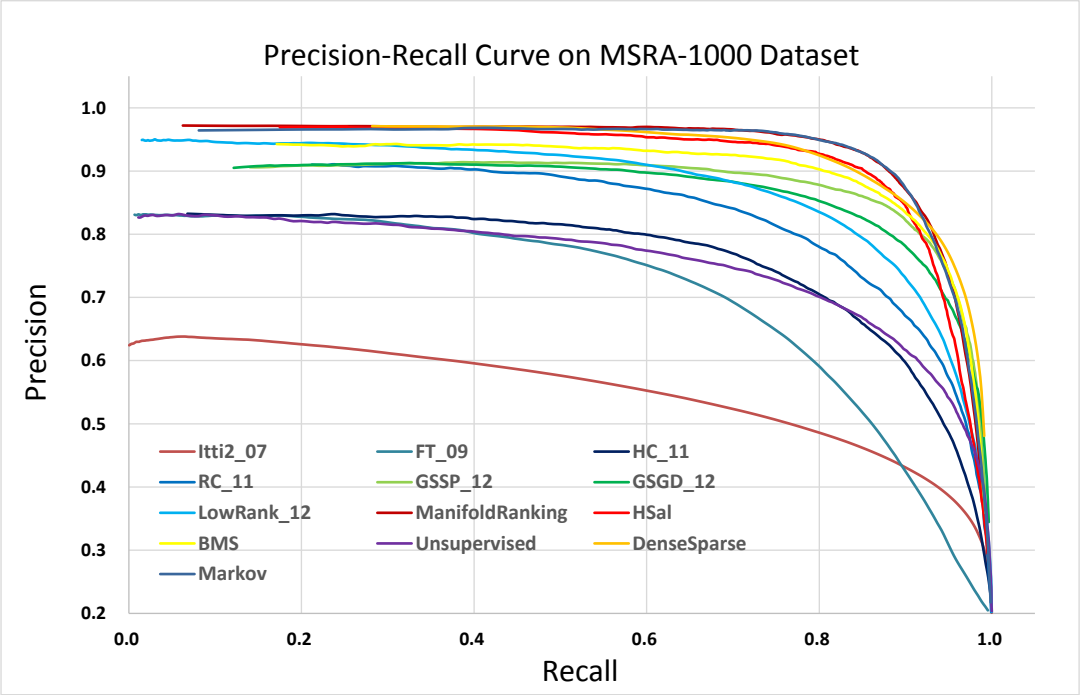


Figure 4: **Average Shuffled-AUC against the STD of Gaussian Blur**. X-axis represents the Gaussian blur standard deviation (STD) in image width and Y-axis represents the average shuffled-AUC score on one dataset.



What exactly is Saliency?

- The term 'Saliency' was first used by Olshausen *et al.* in 1993, *Journal of Neuroscience*.
- Referred to as *Saliency, Visual Attention, Unpredictability,* etc.
- Feature: subjective, ambiguous and task-dependent.
  - Traditionally, where a human looks. (Eye fixation)
  - Recently, where the salient object is. (Salient object detection)
- Solution: biologically based, purely computational, or a combo.

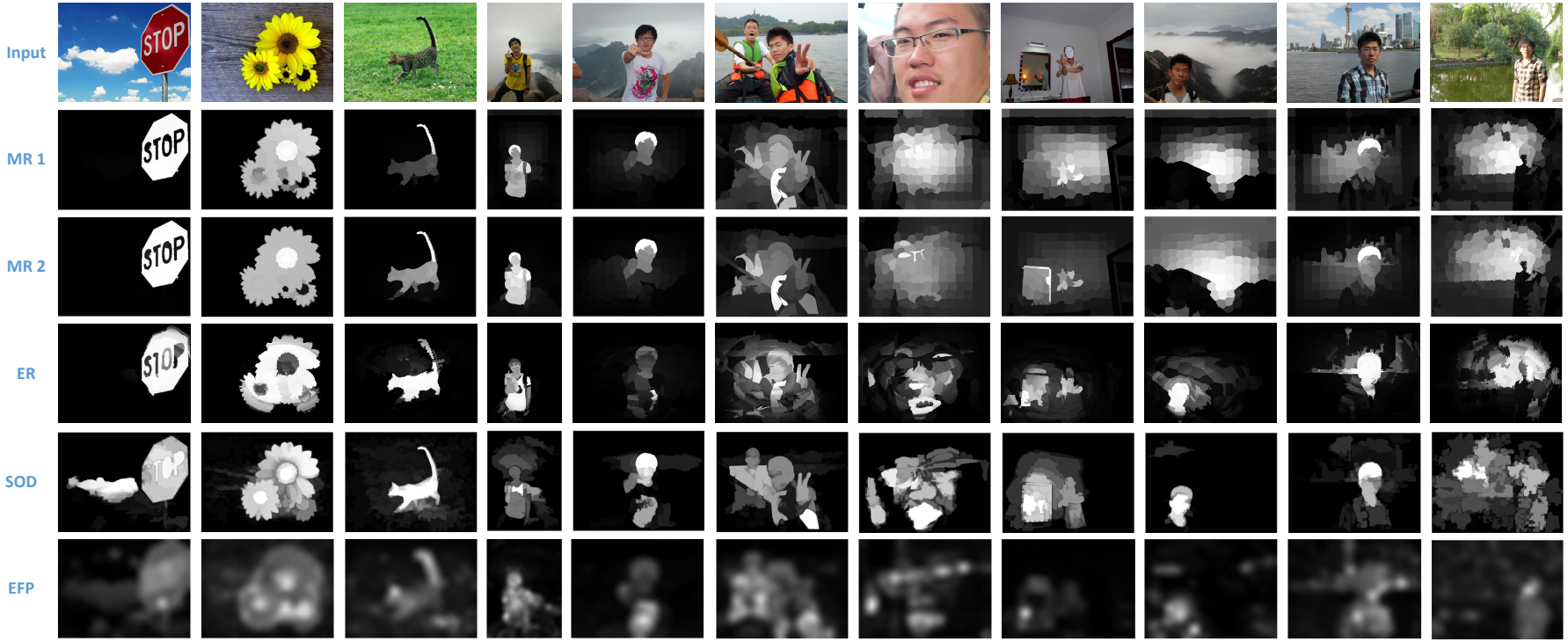
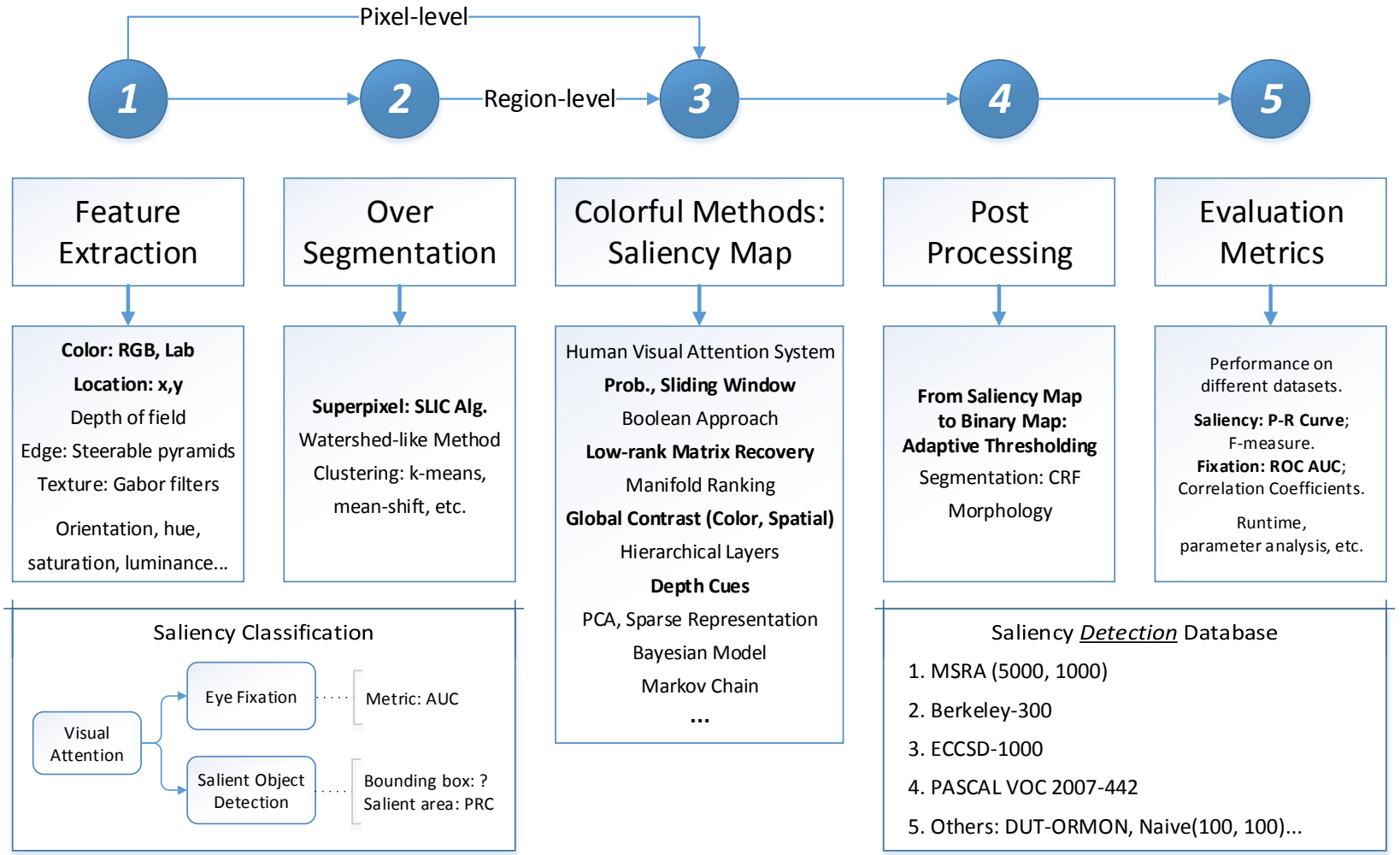


Figure: Performance on Realistic Images. 图片整理：迟至真

Sum Up: General Steps to Do Saliency



The End

Questions and Discussions

Acknowledgement: 刘巍、王建鹏、李晓晖