Salient Region Detection

NINGYI DENG 邓凝旖

2014.9

Ice Breaking

KEY PERSON IN THIS PRESENTATION

Xiaodi Hou 侯晓迪 @ Caltech

An algorithm
- Log spectrum representation of image
- Five-line source code
- [07CVPR] Saliency Detection: A Spectral Residual Approach

A dataset
- PASCAL-S
- [14CVPR]The Secrets of Salient Object Segmentation

A joke
- Two reviews from CVPR reviewer for Xiaodi Hou
  - 5 - Definitely reject This paper is very interesting, and it should definitely be published, but not at NIPS.
  - 3 - References missing Some important references are missing, the followings list a few: Xiaodi Hou, Liqing Zhang: Saliency Detection: A Spectral Residual Approach. CVPR 2007

Outline

- Motivation
- Application
- Two Different Worlds of Saliency
- Perceptual Feature
- Combining Two Tasks
- Problems in Previous Work
- Future Challenges
- Future Directions

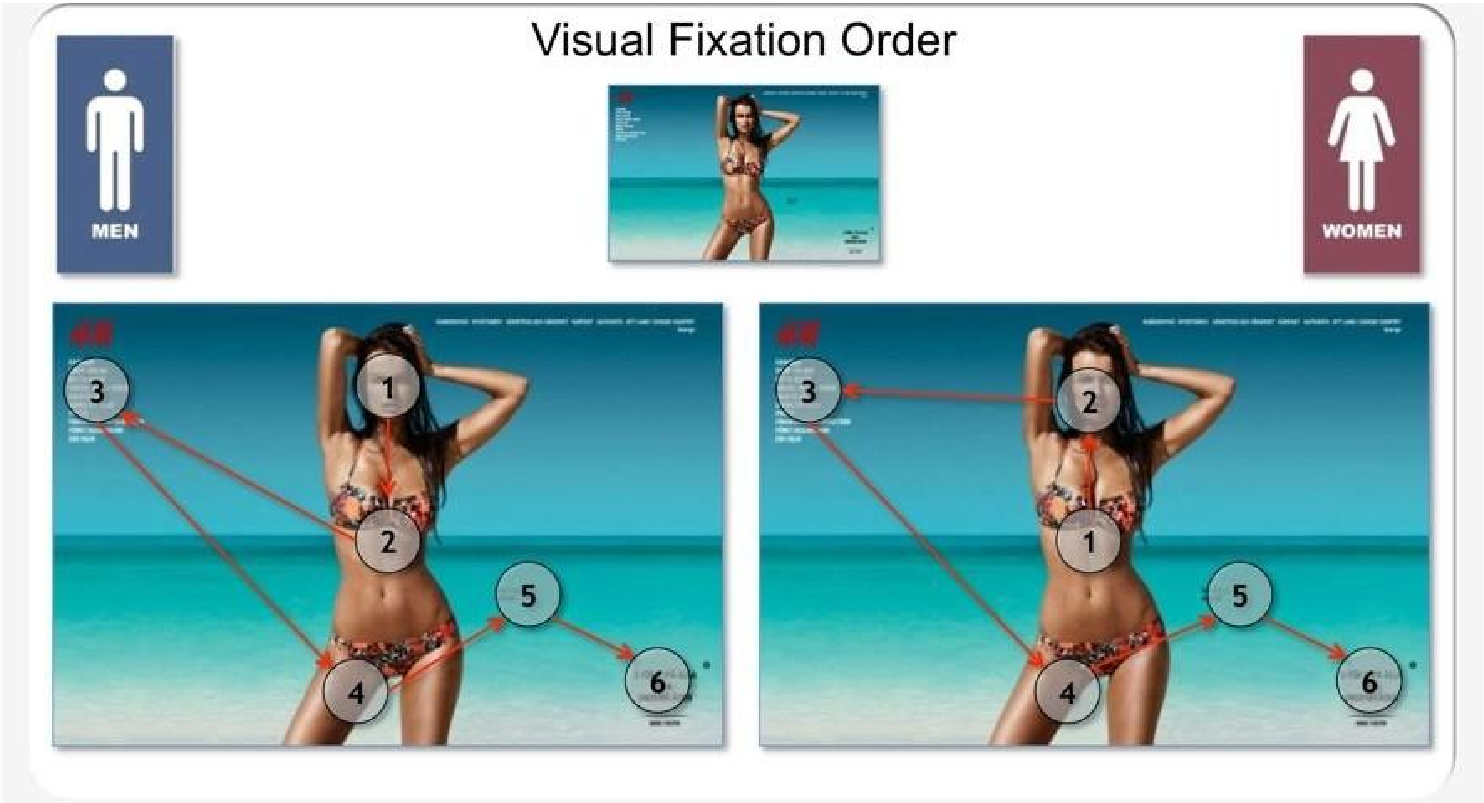
Motivation

How Men And Women Look At Advertisement Differently

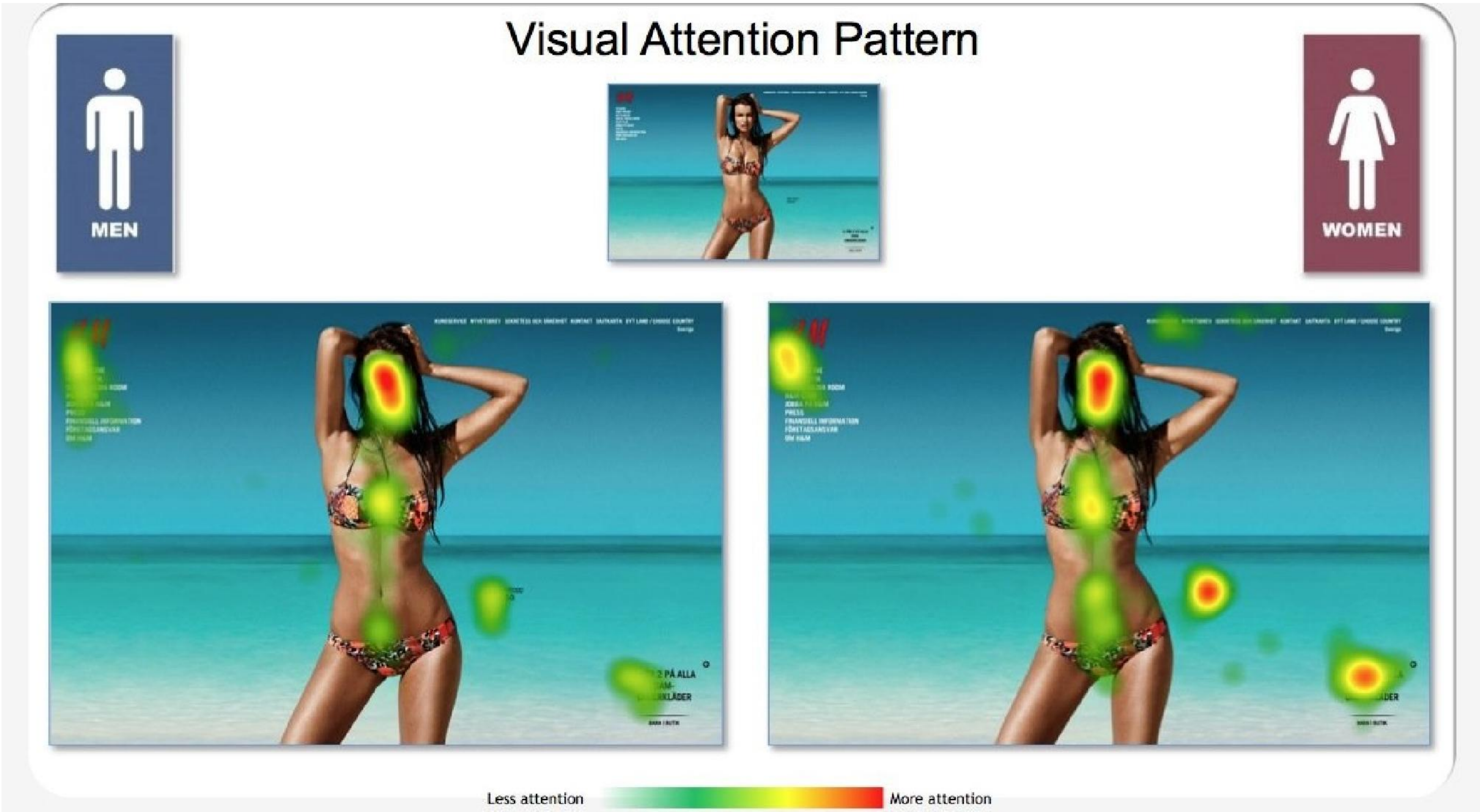
H&M

WEBPAGE

How Men And Women Look At Advertisement Differently



How Men And Women Look At Advertisement Differently



Eye Tracking

Application

CONTENT-AWARE IMAGE/VIDEO RESIZING

OBJECT SEGMENTATION

Content-aware Image/video Resizing

Object Segmentation★

Object Segmentation★

Two Different Tasks of Saliency

FIXATION PREDICTION

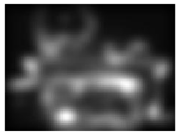
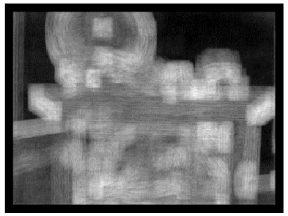
SALIENT OBJECT SEGMENTATION★

Fixation Prediction

Purpose

- To compute a probabilistic map of an image to predict the actual human eye gaze patterns

Result


STATE-OF-THE-ART

- ITTI [Itti et al. PAMI 98]
- AIM [Bruce et al. NIPS 06]
- GBVS [Harel et al. NIPS 07]
- DVA [Hou et al. NIPS 08]
- SUN [Zhang et al. JOV 08]
- SIG [Hou et al. PAMI 12]

Saliency Detection: A Spectral Residual Approach

Log Spectrum Representation

- The amplitude A(f) of the averaged Fourier spectrum of the ensemble of natural images obeys a distribution:

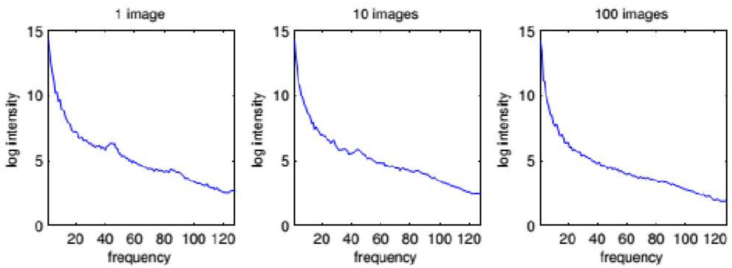


Figure 3. Curves of averaged spectra over 1, 10 and 100 images.

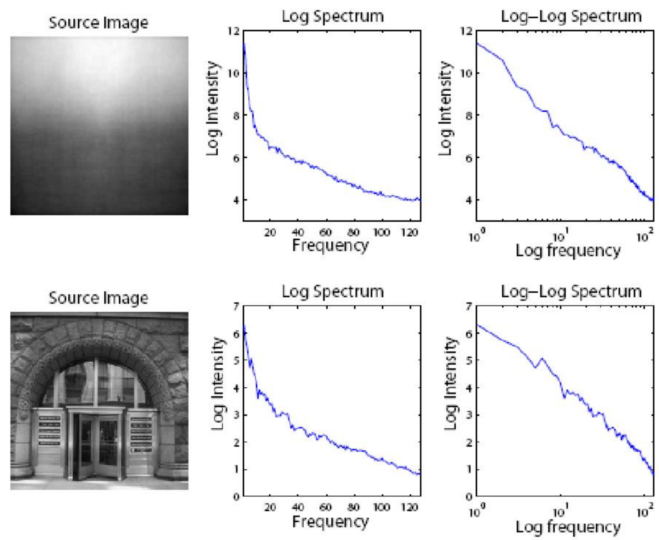


Figure 1. Examples of log spectrum and log-log spectrum. The first image is the average of 2277 natural images.

- Five-line source code

$$
\begin{aligned}
\mathcal{A}(f) &= \Re\Big(\mathfrak{F}\big[\mathcal{I}(x)\big]\Big), \quad \mathcal{P}(f) = \Im\Big(\mathfrak{F}\big[\mathcal{I}(x)\big]\Big), \\
\mathcal{L}(f) &= \log\big(\mathcal{A}(f)\big), \quad \mathcal{R}(f) = \mathcal{L}(f) - h_n(f) * \mathcal{L}(f), \\
\mathcal{S}(x) &= g(x) * \mathfrak{F}^{-1}\Big[\exp\big(\mathcal{R}(f) + \mathcal{P}(f)\big)\Big]^2.
\end{aligned}
$$

Saliency Detection:
A Spectral Residual Approach

Result

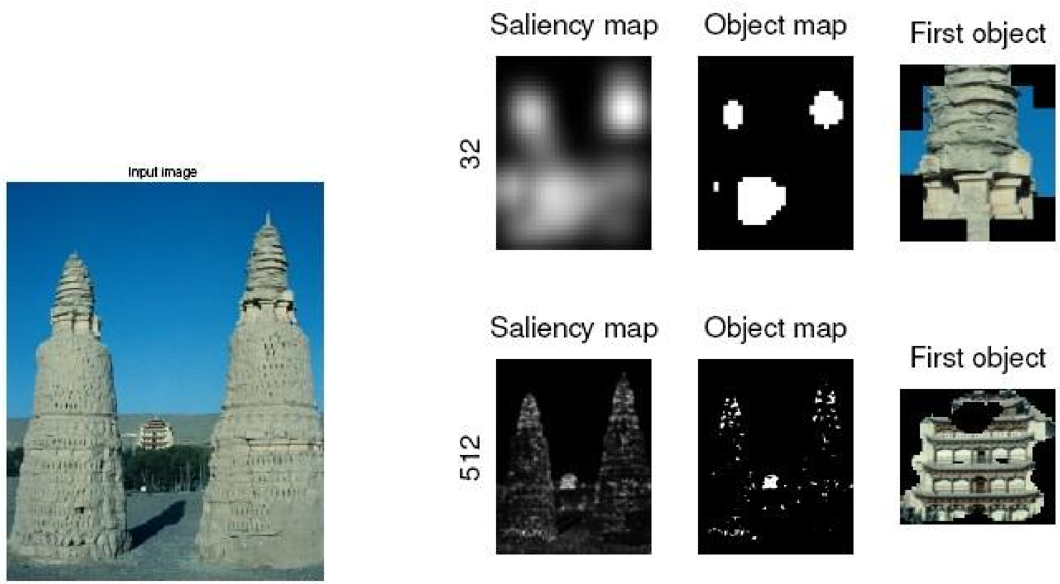


Figure 7. An example of attention in different scales.

Salient Object Segmentation

Purpose

- To generate masks that matches the annotated silhouettes of salient objects

Ground-truth: mask

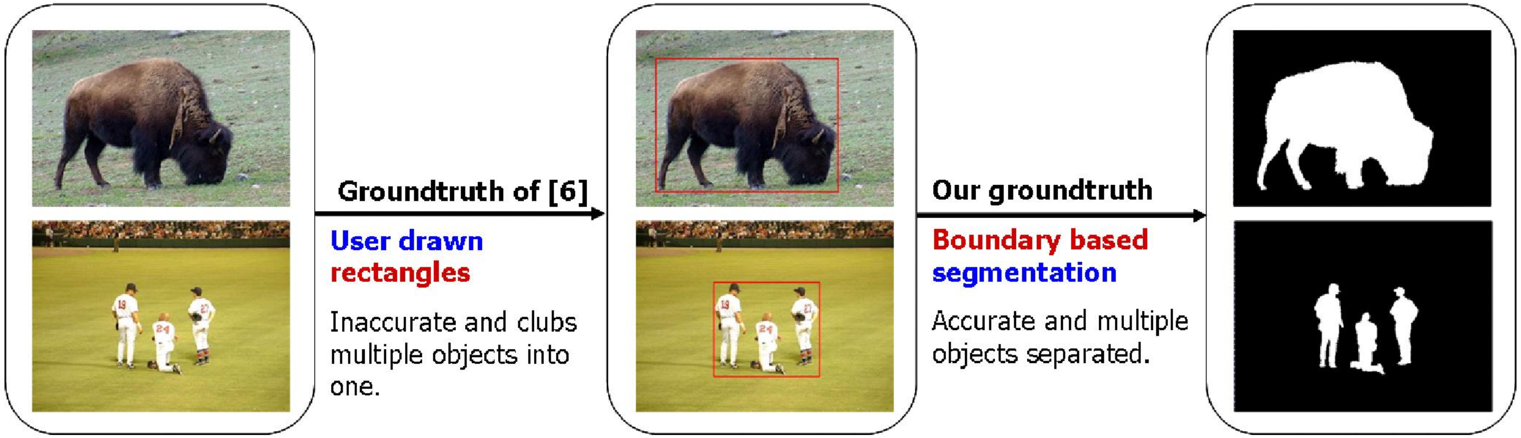
Approaches

- Low-level feature (+ segmentation)
- Popular algorithm
- Perceptual feature

Low-level feature + Segmentation

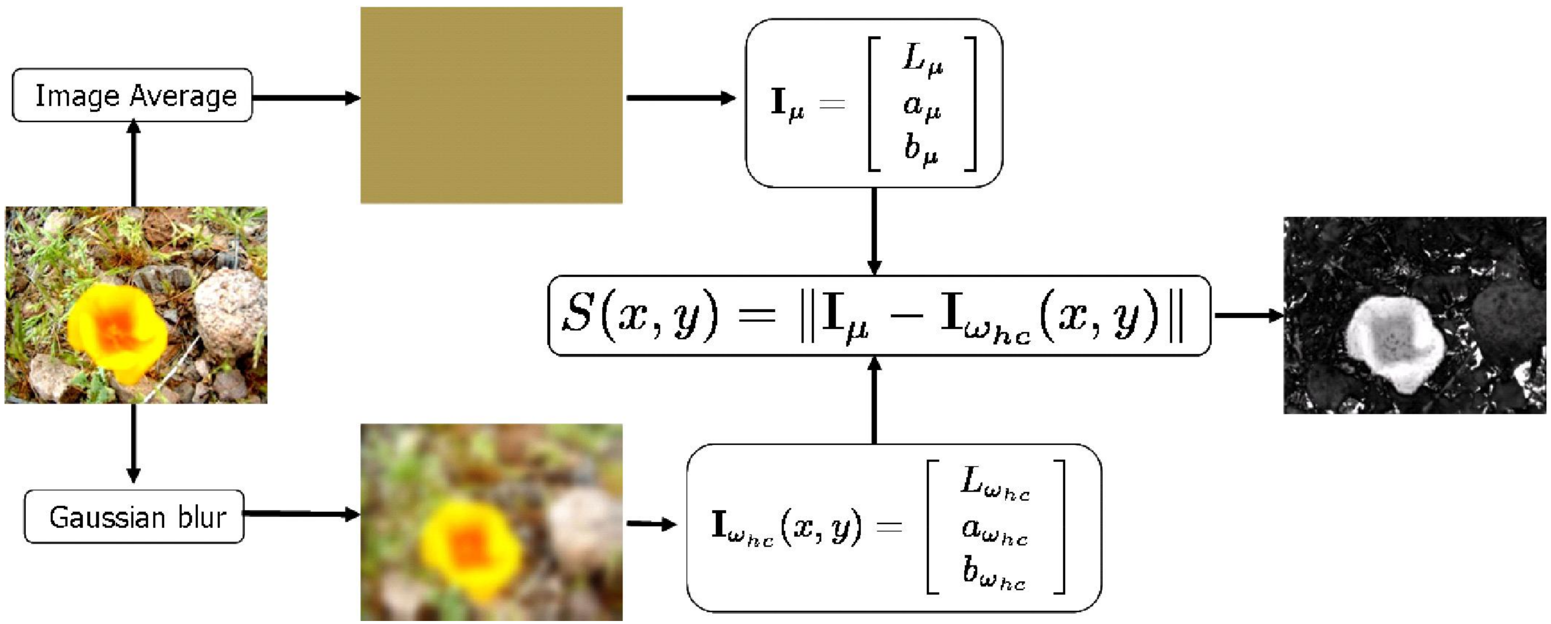
[09CVPR]Frequency-tuned Salient Region Detection

- First using saliency for outstand region detection
- Build up first dataset with mask of objects



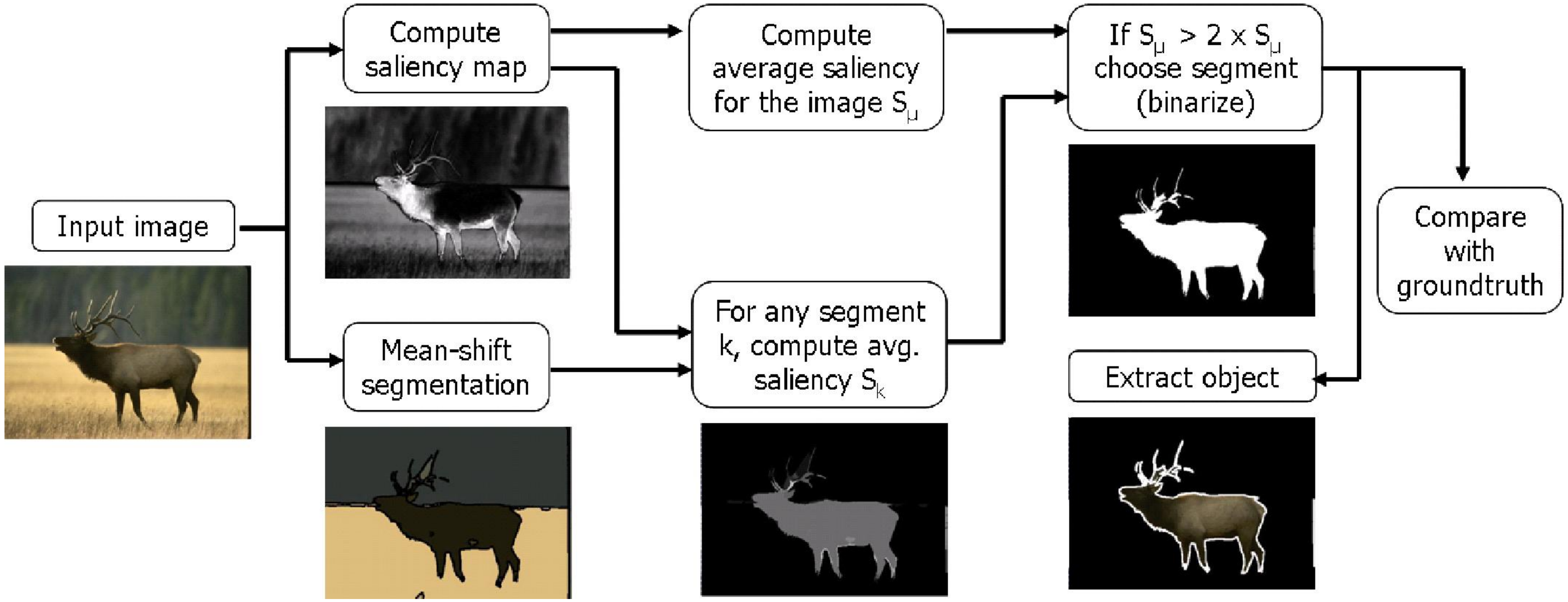
FT- Low-level feature

Color & brightness



FT-Saliency + Segmentation

Combine with mean-shift



FT-Result

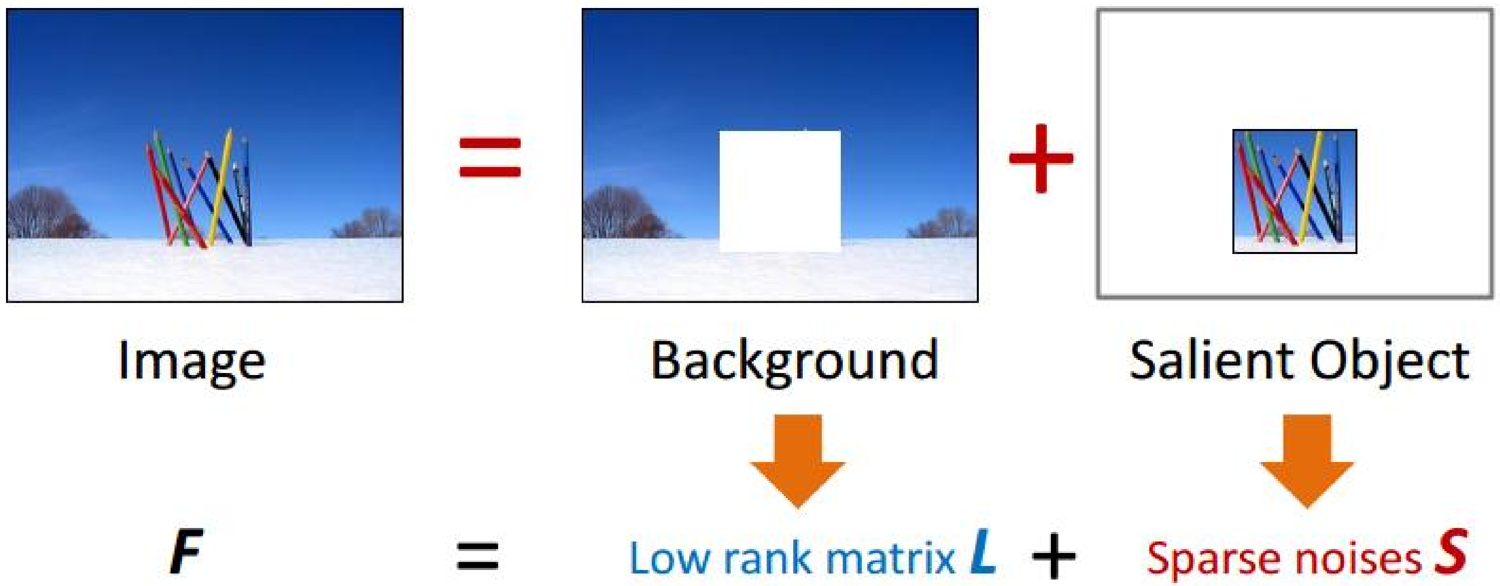
Real image for test

Result

Popular algorithm

[12CVPR]A Unfied Approach to Salient Object Detection via Low Rank Matrix Recovery

- Low rank matrix
- Higher-level prior integration



LR-Simplifying Low-level Matrix by Pre-Processing

$$(\mathbf{L}^*, \mathbf{S}^*) = arg \min_{\mathbf{L},\mathbf{S}} (rank(\mathbf{L}) + \lambda \| \mathbf{S} \|_0)$$
$$s.t. \quad \mathbf{F} = \mathbf{L} + \mathbf{S}$$

NP Hard...

Unified Model

$$(\mathbf{L}^*, \mathbf{S}^*) = arg \min_{\mathbf{L},\mathbf{S}} (\| \mathbf{L} \|_* + \lambda \| \mathbf{S} \|_1)$$
$$s.t. \; \mathbf{TFP} = \mathbf{L} + \mathbf{S}$$

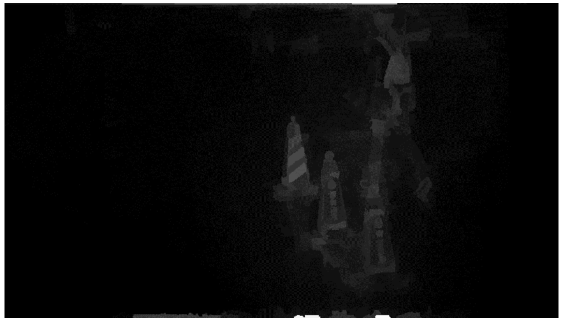
T : a learned linear metric

P : a higher-level prior map

LR-Result

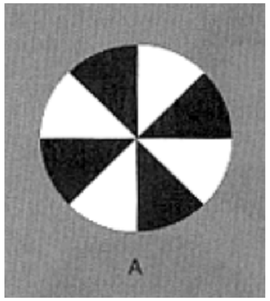
Real image for test

Result

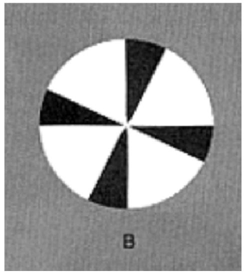


Perceptual Feature

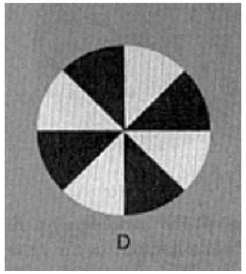
Figure/Ground Organization in visual system



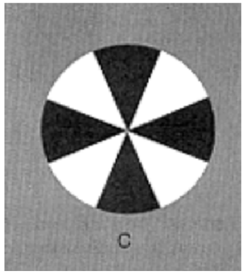
Surroundedness



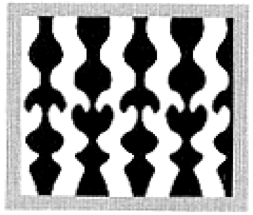
Size



Contrast

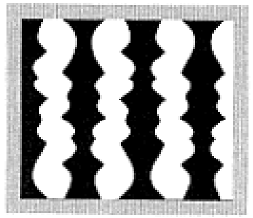


Horizontal/vertical orientation



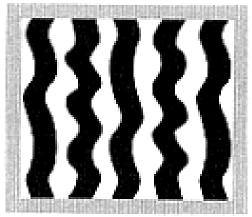
A

Symmetry



B

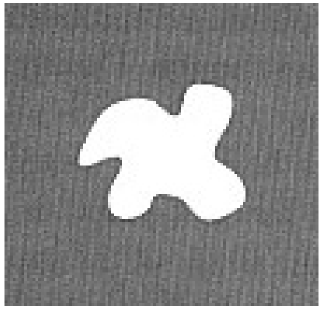
Convexity



C

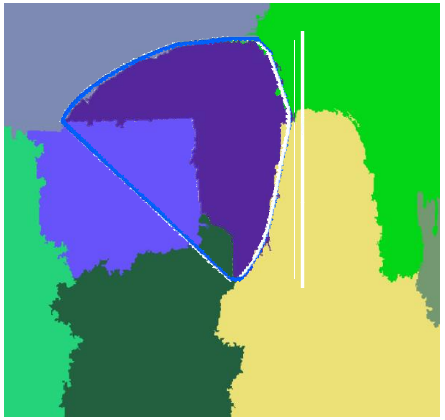
Parallelism

Figure/Ground Organization in visual system

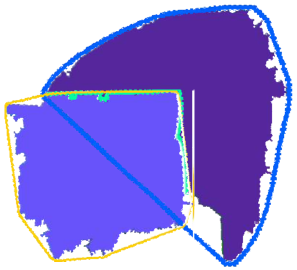


Convexity

Surroundedness



$$convex(r_i) = 1 - \frac{size(CH_i) - size(r_i)}{size(CH_i)}$$



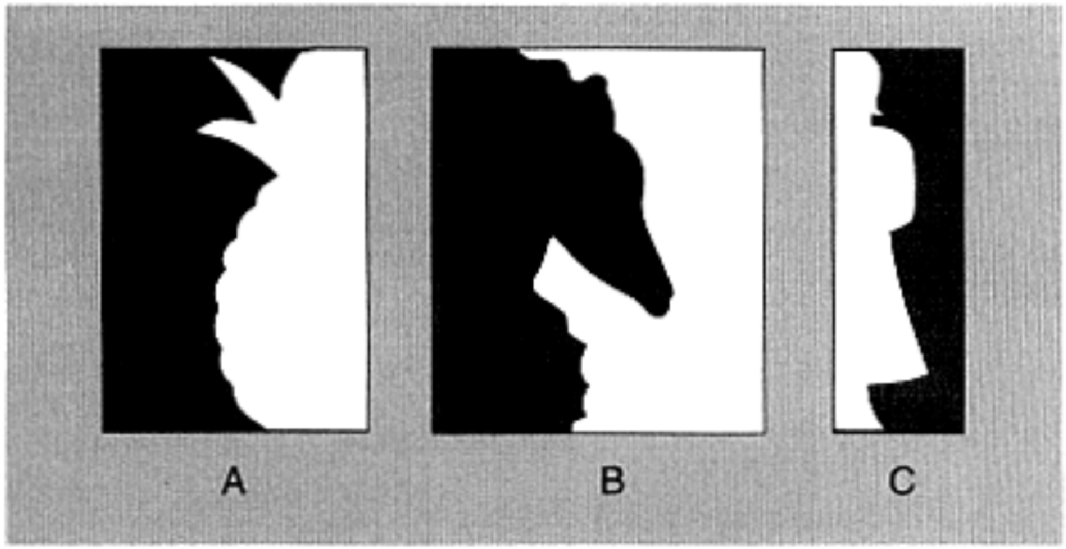
$$surround(r_i) = \frac{\max_{r_j \in S_i}\left(C(r_i) \cap C(r_j)\right)}{C(r_i)}$$

$$surround*(r_i) = \frac{\max_{r_j \in S_i}\left(size(CH_i) \cap size(CH_j)\right)}{\mathrm{size}(CH_i)}$$

Figure/Ground Organization in visual system

Meaningfulness

- High-level features
- Top-down info, i.e. Classification



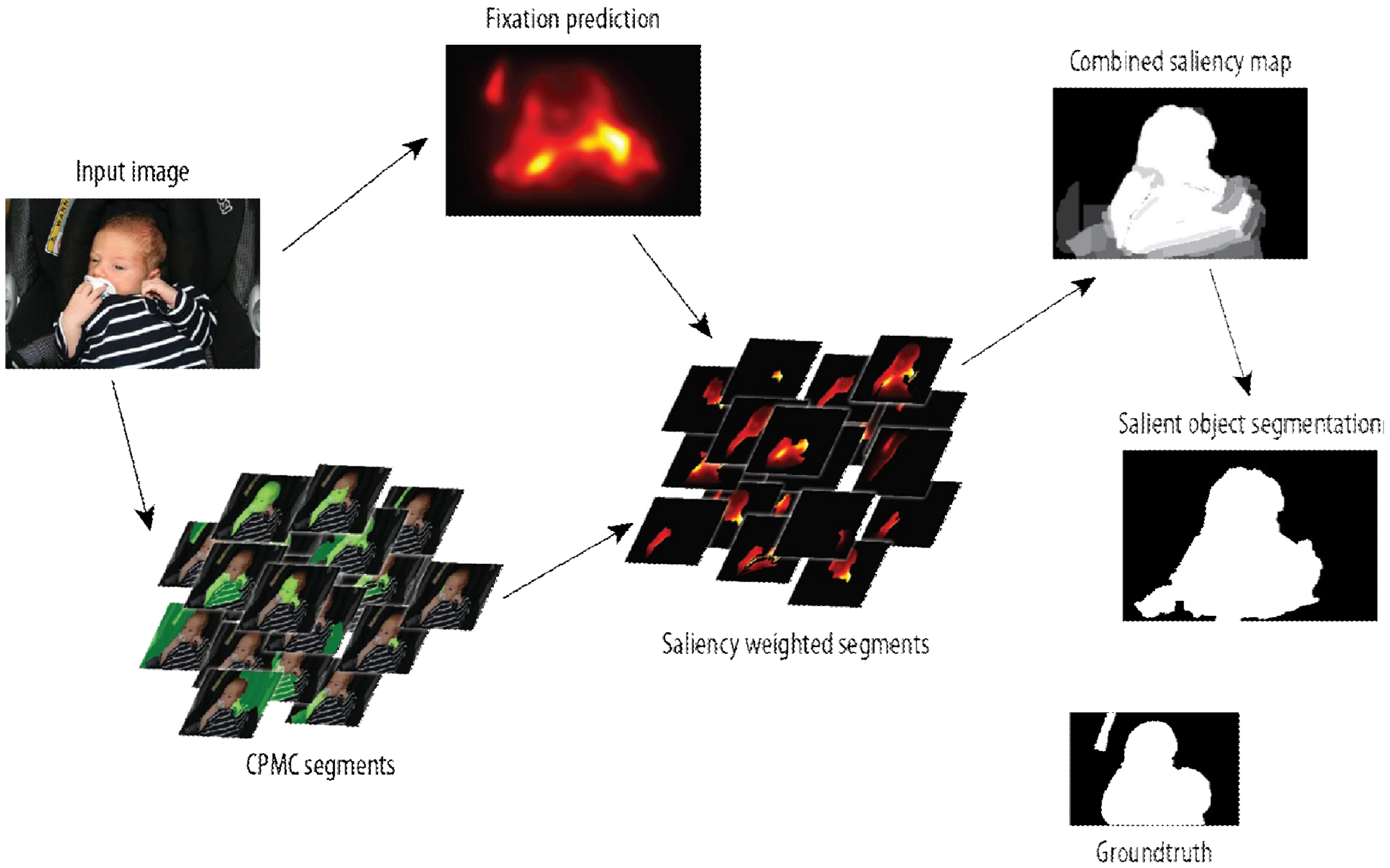
Combining Two Tasks

FIXATION PREDICTION + SEGMENTATION

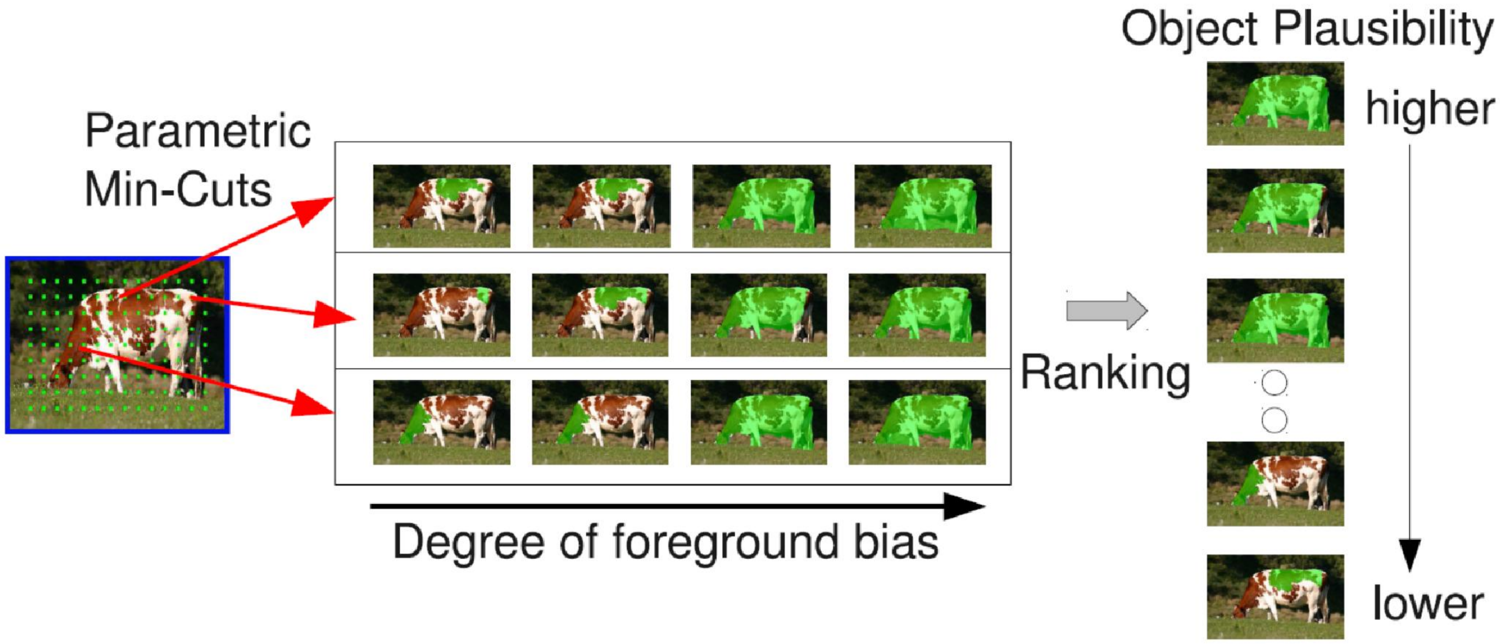
≈ SALIENT OBJECT SEGMENTATION

FROM *THE SECRETS OF SALIENT OBJECT SEGMENTATION*

Combination Model

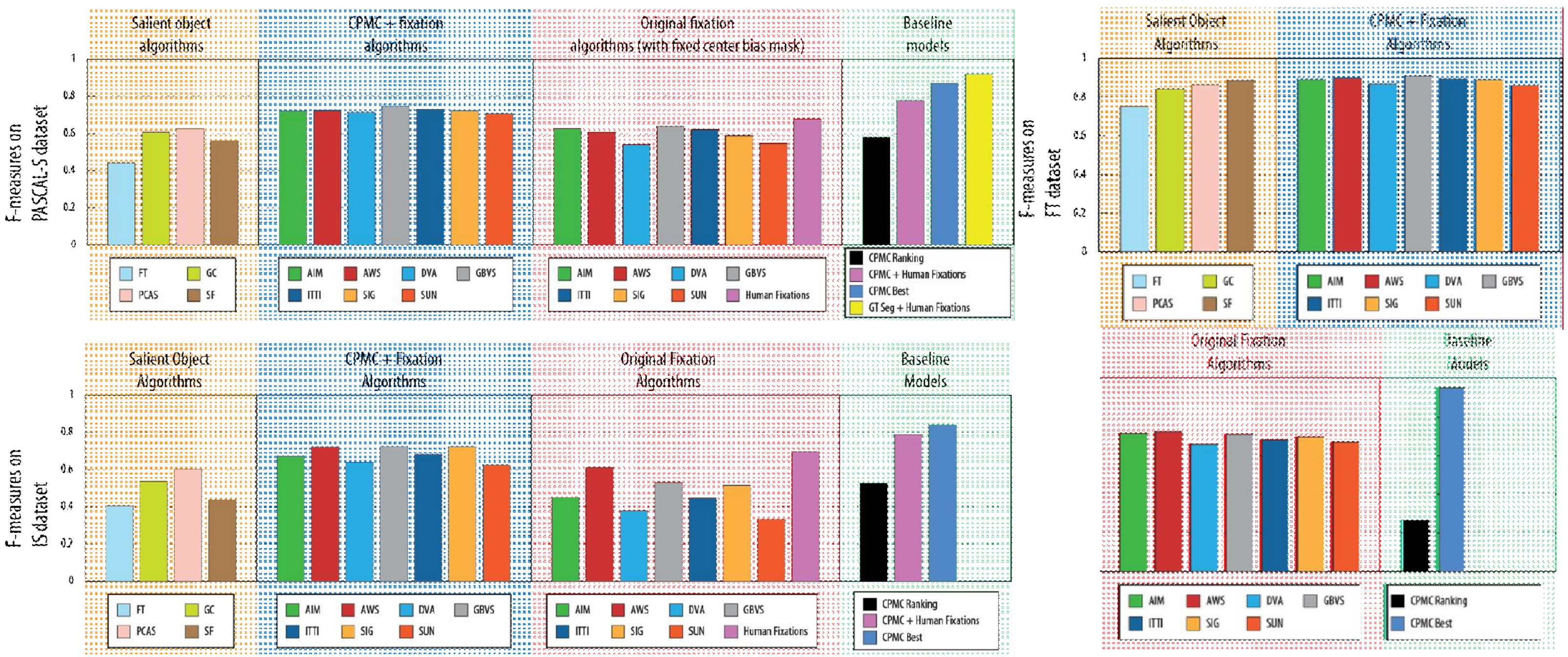


CMPC



Object Plausibility
higher
Parametric
Min-Cuts
Ranking
Degree of foreground bias
lower

Result



Salient object algorithms
CPMC + Fixation algorithms
Original fixation algorithms (with fixed center bias mask)
Baseline models
F-measures on PASCAL-S dataset
FT
GC
PCAS
SF
AIM
AWS
DVA
GBVS
ITTI
SIG
SUN
Human Fixations
CPMC Ranking
CPMC + Human Fixations
CPMC Best
GT Seg + Human Fixations
Salient Object Algorithms
CPMC + Fixation Algorithms
Original Fixation Algorithms
Baseline Models
F-measures on IS dataset
F-measures on FT dataset

Problems in Previous Work

ALSO FROM *THE SECRETS OF SALIENT OBJECT SEGMENTATION*

Problems

Dataset bias

- Salient object segmentation dataset is heavily biased
- Images in FT dataset usually have a foreground object with discernable boundaries being
- Surrounded by background that have contrastive colors

Center-bias

- Easily applied by simple function, e.g. Gaussian filter
- Do great contribution to saliency (human fixations ***are*** heavily biased towards the center of the image

Future Directions

IMAGE DETECTION

VIDEO DETECTION

COMBINATION WITH OBJECTNESS

Future Challenges

Image detection

- Segmentation
- Precision

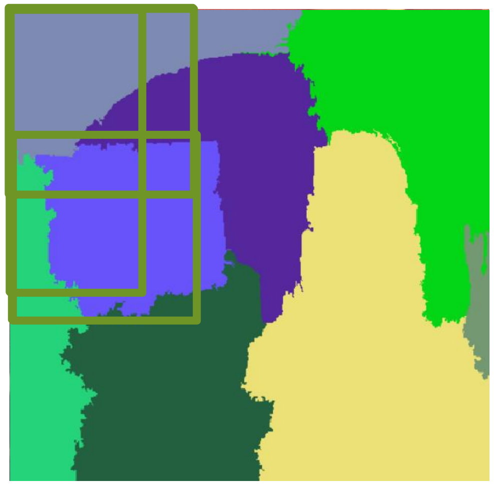
Video detection

- Time
- Motion information

Future Directions

Combination with Objectness

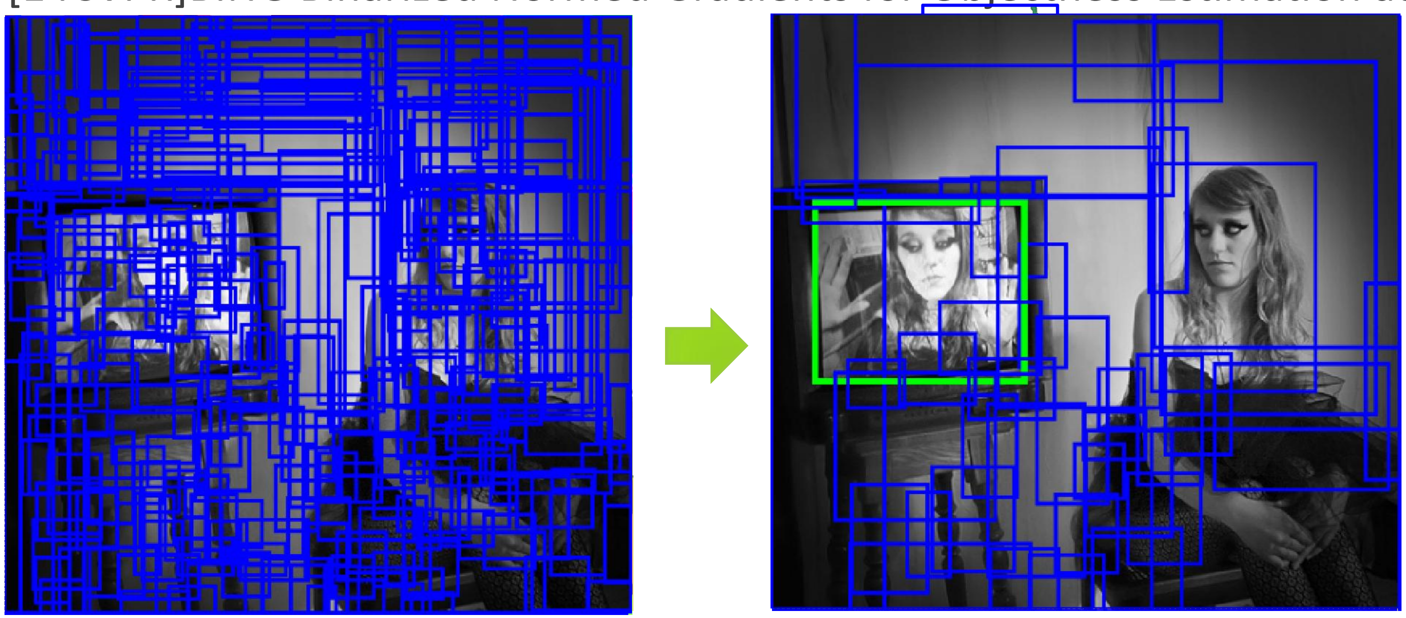
- **[10CVPR]What is an object**
- [13IJCV]Selective Search for Object Recognition
- [14CVPR]BING Binarized Normed Gradients for Objectness Estimation at 300fps



Future Directions

Combination with Objectness

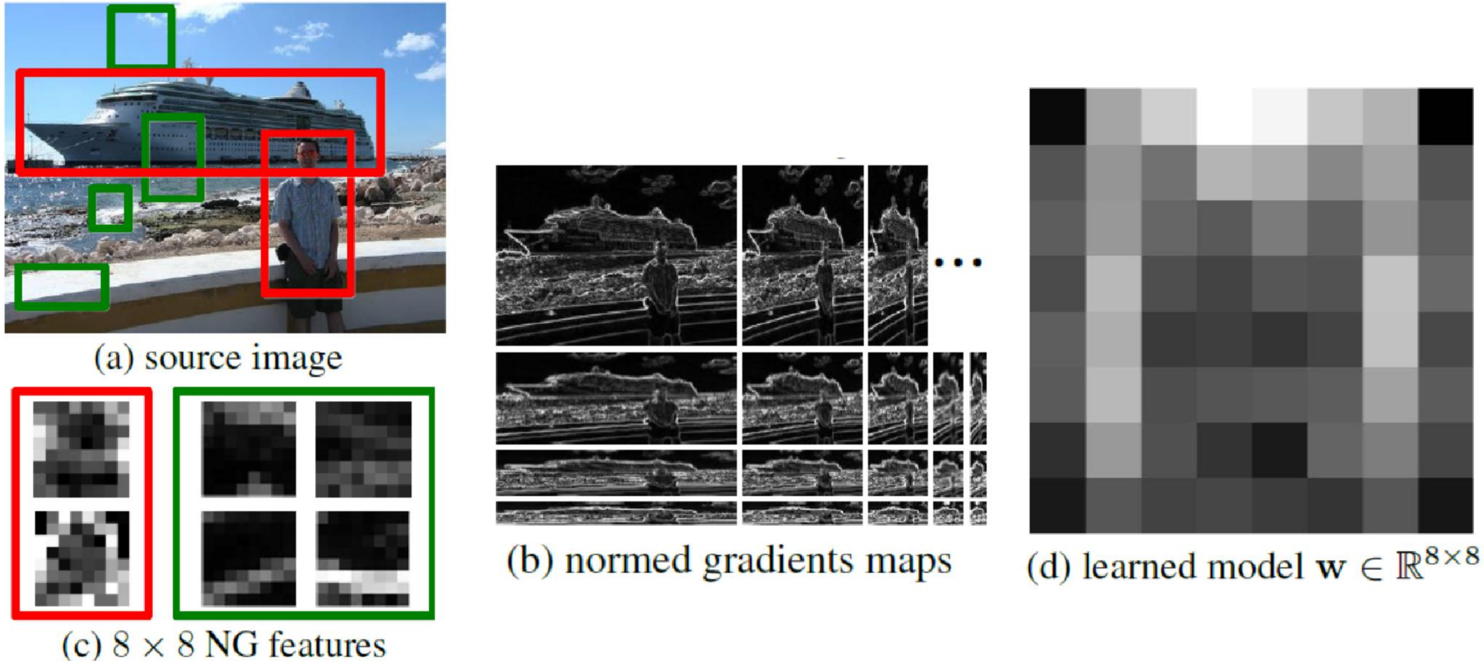
- [10CVPR]What is an object
- **[13IJCV]Selective Search for Object Recognition**
- [14CVPR]BING Binarized Normed Gradients for Objectness Estimation at



Future Directions

Combination with Objectness

- [10CVPR]What is an object
- [13IJCV]Selective Search for Object Recognition
- **[14CVPR]BING Binarized Normed Gradients for Objectness Estimation**



[10CVPR]What is an object

Image cues

Multi-scale Saliency (MS)

$$\mathrm{MS}(w,\theta^s_{\mathrm{MS}}) = \sum_{\{p\in w \mid I^s_{\mathrm{MS}}(p)\geq\theta_s\}} I^s_{\mathrm{MS}}(p) \times \frac{|\{p\in w \mid I^s_{\mathrm{MS}}(p)\geq\theta_s\}|}{|w|}$$

Color Contrast (CC)

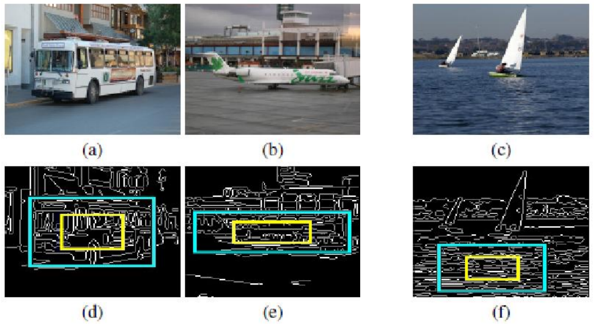
$$\mathrm{CC}(w,\theta_{CC}) = \chi^2(h(w),\ h(\mathrm{Surr}(w,\theta_{CC})))$$

$$\frac{|\mathrm{Surr}(w,\theta_{CC})|}{|w|} = \theta^2_{CC} - 1$$

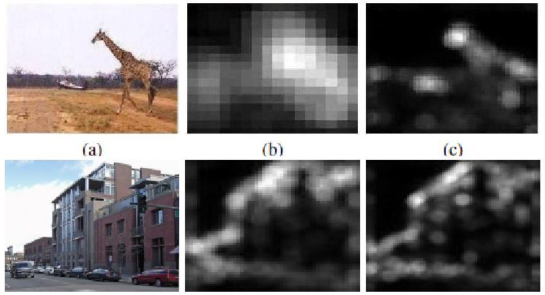
Edge Density (ED)

$$\mathrm{ED}(w,\theta_{ED}) = \frac{\sum_{p\in\mathrm{Inn}(w,\theta_{ED})} I_{ED}(p)}{\mathrm{Len}(\mathrm{Inn}(w,\theta_{ED}))}$$

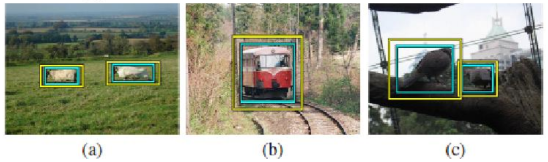
$$\frac{|\mathrm{Inn}(w,\theta_{ED})|}{|w|} = 1/\theta^2_{ED}$$



Canny Calculator

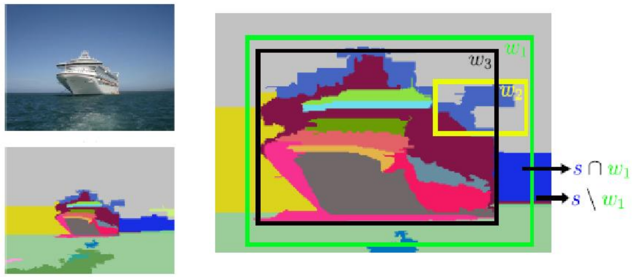


Saliency Detection:
A Spectral Residual Approach



Superpixels Straddling (SS)

$$\mathrm{SS}(w,\theta_{SS}) = 1 - \sum_{s\in S(\theta_{SS})} \frac{\min(|s\setminus w|, |s\bigcap w|)}{|w|}$$



[13IJCV]Selective Search for Object Recognition

Similarity Measures

Colour

$$s_{colour}(r_i, r_j) = \sum_{k=1}^{n} \min(c_i^k, c_j^k)$$

Texture

$$s_{texture}(r_i, r_j) = \sum_{k=1}^{n} \min(t_i^k, t_j^k)$$

Size

$$s_{size}(r_i, r_j) = 1 - \frac{\text{size}(r_i) + \text{size}(\text{r}_\text{j})}{\text{size}(im)}$$

Fill

$$fill(r_i, r_j) = 1 - \frac{\text{size}(BB_{ij}) - \text{size}(r_i) - \text{size}(r_i)}{\text{size}(im)}$$

[14CVPR]BING Binarized Normed Gradients for Objectness Estimation

For linear model w:

$$\mathbf{w} \approx \sum_{j=1}^{N_w} \beta_j \mathbf{a}_j$$

$$\mathbf{a}_j \in \{-1, 1\}^{64}$$

$$\mathbf{a}_j = \mathbf{a}_j^+ - \overline{\mathbf{a}_j^+}$$

$$\mathbf{a}_j^+ \in \{0, 1\}^{64}$$

$$\langle \mathbf{w}, \mathbf{b} \rangle \approx \sum_{j=1}^{N_w} \beta_j (2\langle \mathbf{a}_j^+, \mathbf{b} \rangle - |\mathbf{b}|)$$

For normed gradients feature NG: using the top Ngbinary bits of the BYTE values

$$\mathbf{g}_l = \sum_{k=1}^{N_g} 2^{8-k} \mathbf{b}_{k,l}$$

For SVM compute:

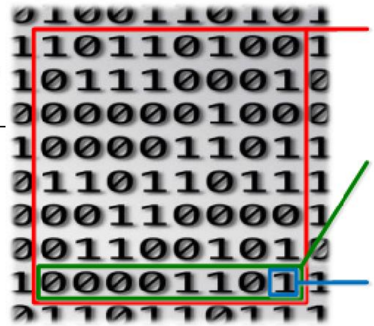
$$s_l \approx \sum_{j=1}^{N_w} \beta_j \sum_{k=1}^{N_g} C_{j,k}$$

For window shift:

$$C_{j,k} = 2^{8-k}(2\langle \mathbf{a}_j^+, \mathbf{b}_{k,l} \rangle - |\mathbf{b}_{k,l}|)$$

Algorithm 2 Get BING features for $W \times H$ positions.

Comments: see Fig. 2 for illustration of variables
Input: binary normed gradient map $b_{W \times H}$
Output: BING feature matrix $\mathbf{b}_{W \times H}$
Initialize: $\mathbf{b}_{W \times H} = 0, \mathbf{r}_{W \times H} = 0$
for each position (x, y) in scan-line order **do**
  $\mathbf{r}_{x,y} = (\mathbf{r}_{x-1,y} \ll 1) \mid b_{x,y}$
  $\mathbf{b}_{x,y} = (\mathbf{b}_{x,y-1} \ll 8) \mid \mathbf{r}_{x,y}$
end for



Future Directions

Object proposal

- Superpixel feature
- Similarity measure
- Re-ranking

Combining top-down information

- Preprocess image for semantic segmentation or classification
- Get top-down info from these following processing

Thanks!
Q&A

Reference

Research from MRC: How do men and women look at advertisements differently?

Y. Li, X. Hou, and C. Koch. The Secrets of Salient Object Segmentation. In CVPR, 2014.

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

R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In CVPR, 2009.

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

M. Sun, A. Farhadi, B. Taskar, S. Seitz. Salient Montages from Unconstrained Videos. In ECCV, 2014

B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In CVPR, 2010.

J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013.

M.-M Cheng, Z. Zhang, W.-Y. Lin, and P. Torr. BING: Binarized Normed Gradients for Objectness Estimation at 300fps . In CVPR 2014.

Palmer, S.E. Vision Science: Photons to Phenomenology. MIT Press. ISBN 978-0-262-16183-1. 1999.