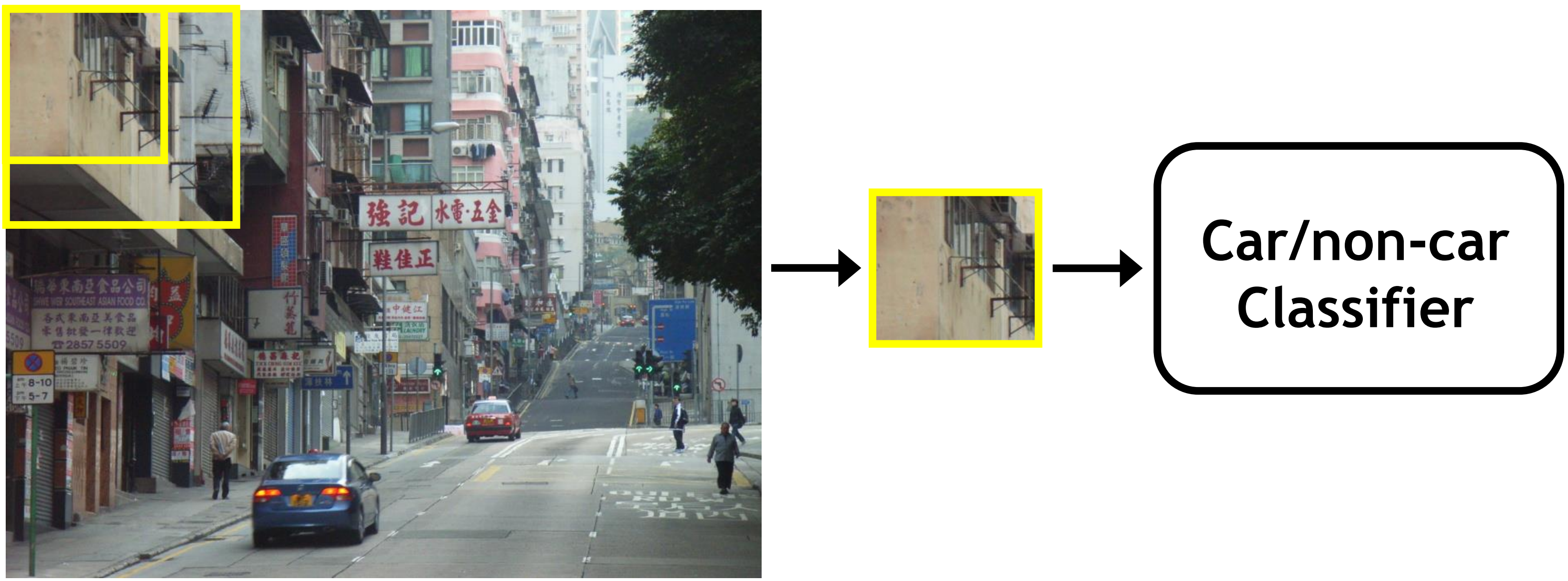
目标检测与识别

Object Detection and Recognition

物体检测

在复杂背景下，通过滑动窗口（sliding windows）搜索感兴趣的物体。



视觉显著性与似物性采样

内容

- 什么是视觉显著性
- 怎样提取显著性区域
- 似物性采样

什么是显著图？

Visual Attention Pattern

Less attention — More attention

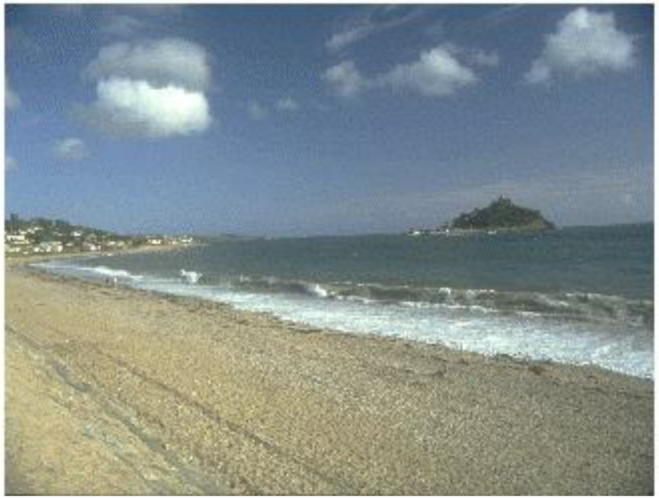
EYETRACKSHOP

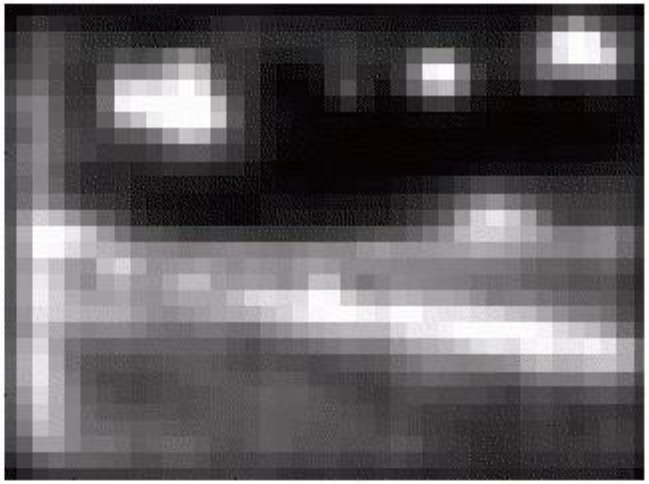
Visual Evaluation

Visual Fixation Order

What is Saliency map?

- The **Saliency Map** is a topographically arranged map that represents visual saliency of a corresponding visual scene.





How to detect Saliency map?

- Buttom-up approach
  - L. Itti’s approach
  - Spectral Residual approach
  - Frequency-tuned approach
  - Global contrast based approach

- Top-down approach
  - Context-aware

How to detect Saliency map?

- Button-up approach
  - L. Itti's approach
  - Spectral Residual approach
  - Frequency-tuned approach
  - Global contrast based approach

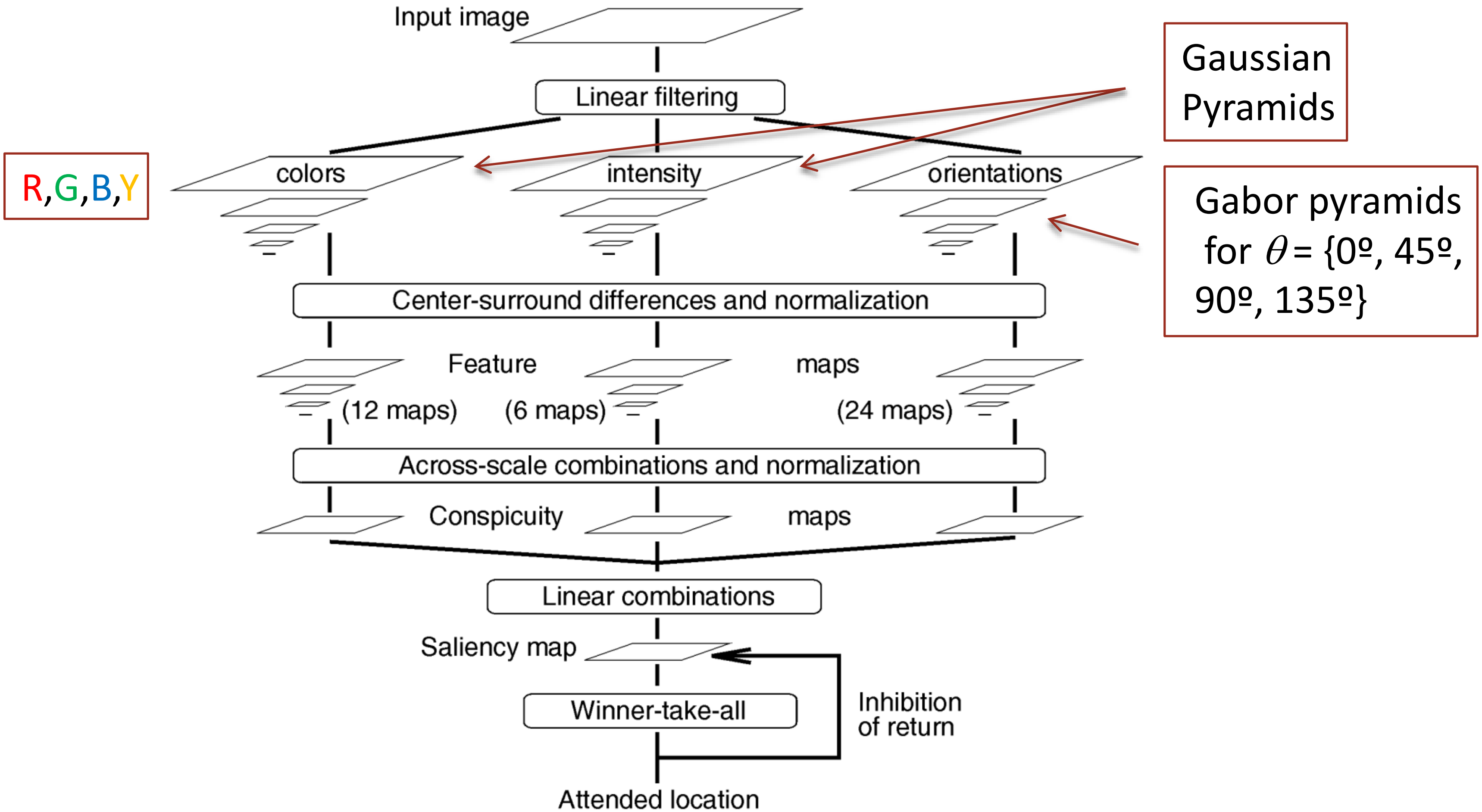
- Top-down approach
  - Context-aware

How to detect Saliency map?

- Button-up approach
  - L. Itti’s approach
  - Spectral Residual approach
  - Frequency-tuned approach
  - Global contrast based approach

- Top-down approach
  - Context-aware

L. Itti's approach (TPAMI,1998)

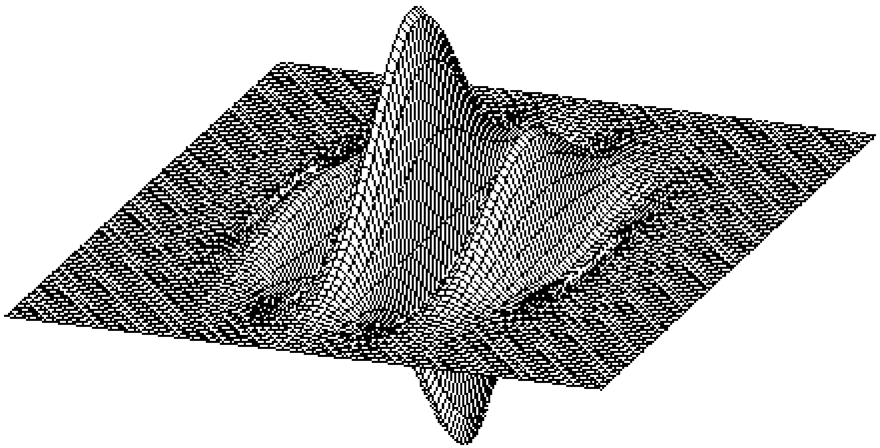


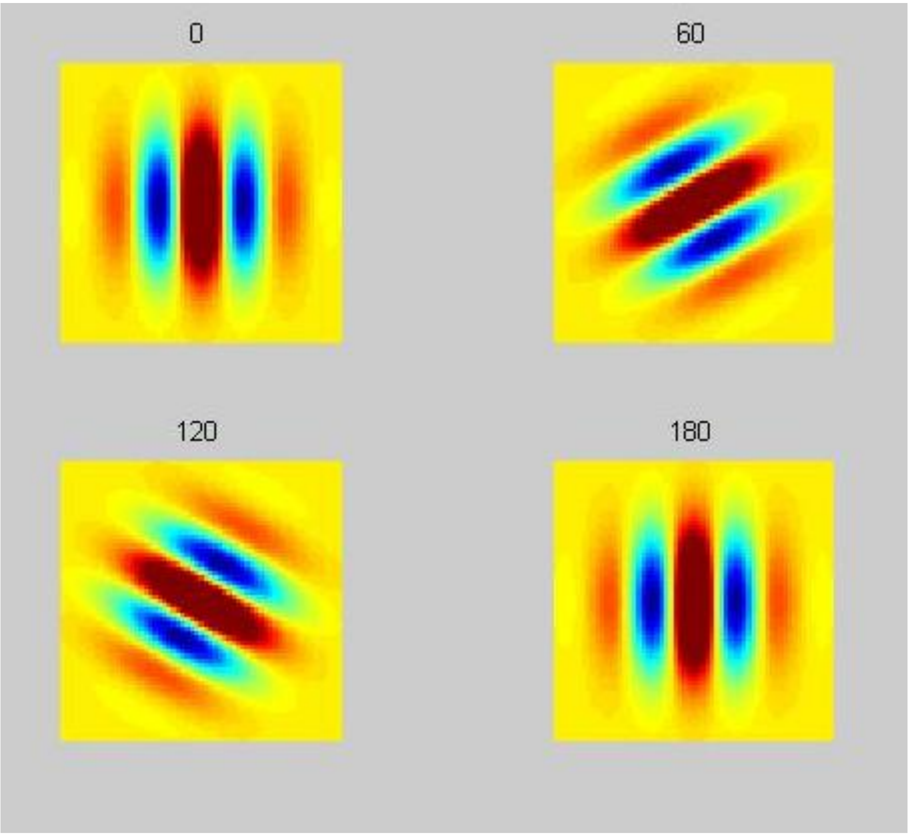
Gaussian Pyramid

Gabor Filter

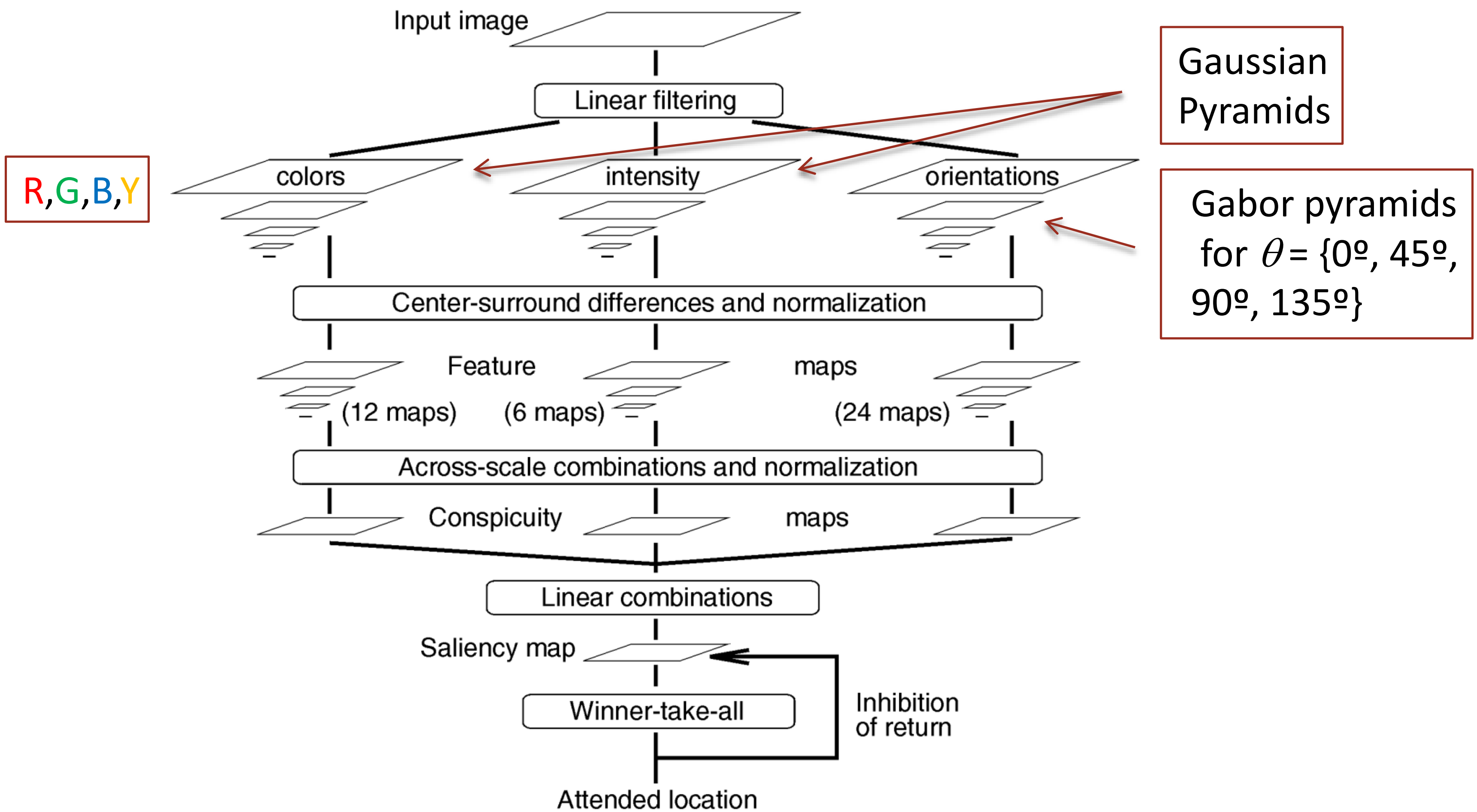
- **Gabor filter**, is a linear filter used for edge detection, texture representation and discrimination.
- In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.
- J. G. Daugman discovered that simple cells in the visual cortex of mammalian brains can be modeled by Gabor functions. Thus, image analysis by the Gabor functions is similar to perception in the human visual system.

Gabor Filter



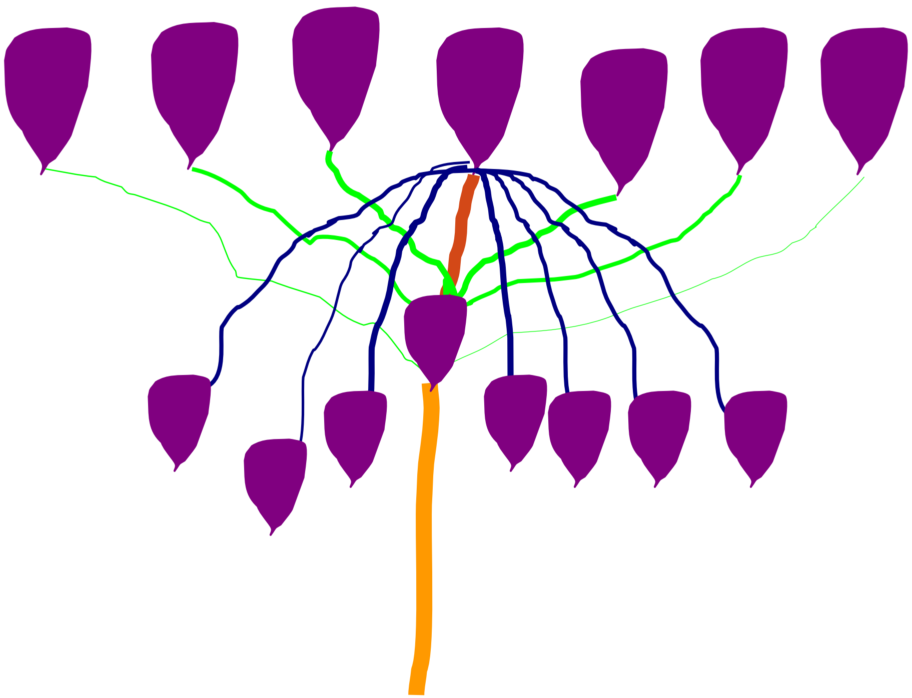


L. Itti's approach (TPAMI,1998)



视觉神经结构

- 感受野:直接或间接影响某一特定神经细胞的光感受器细胞的全体



同心圆感受野

- 人的视觉细胞存在视觉场结构.视点的中心区域存在正性细胞. 它们接收光能并产生一个正的反应。在该中心区域周围存在着负性细胞. 它们在接收光能时产生相反的反应。负性细胞随中心距增大而迅速稀疏，代之而起的中性细胞不产生任何反应。这种解释由诺贝尔奖金获得者Hartline得到证实。
- 这种场结构所产生的视觉反应可由”墨西哥草帽”来表示.
- 这种场结构可以使人的视觉具有侧抑制作用，它使观察物体时保证“集中注意力”. 即把视觉活动集中在注意圈内，不受圈外的变化所干扰。

同心圆感受野

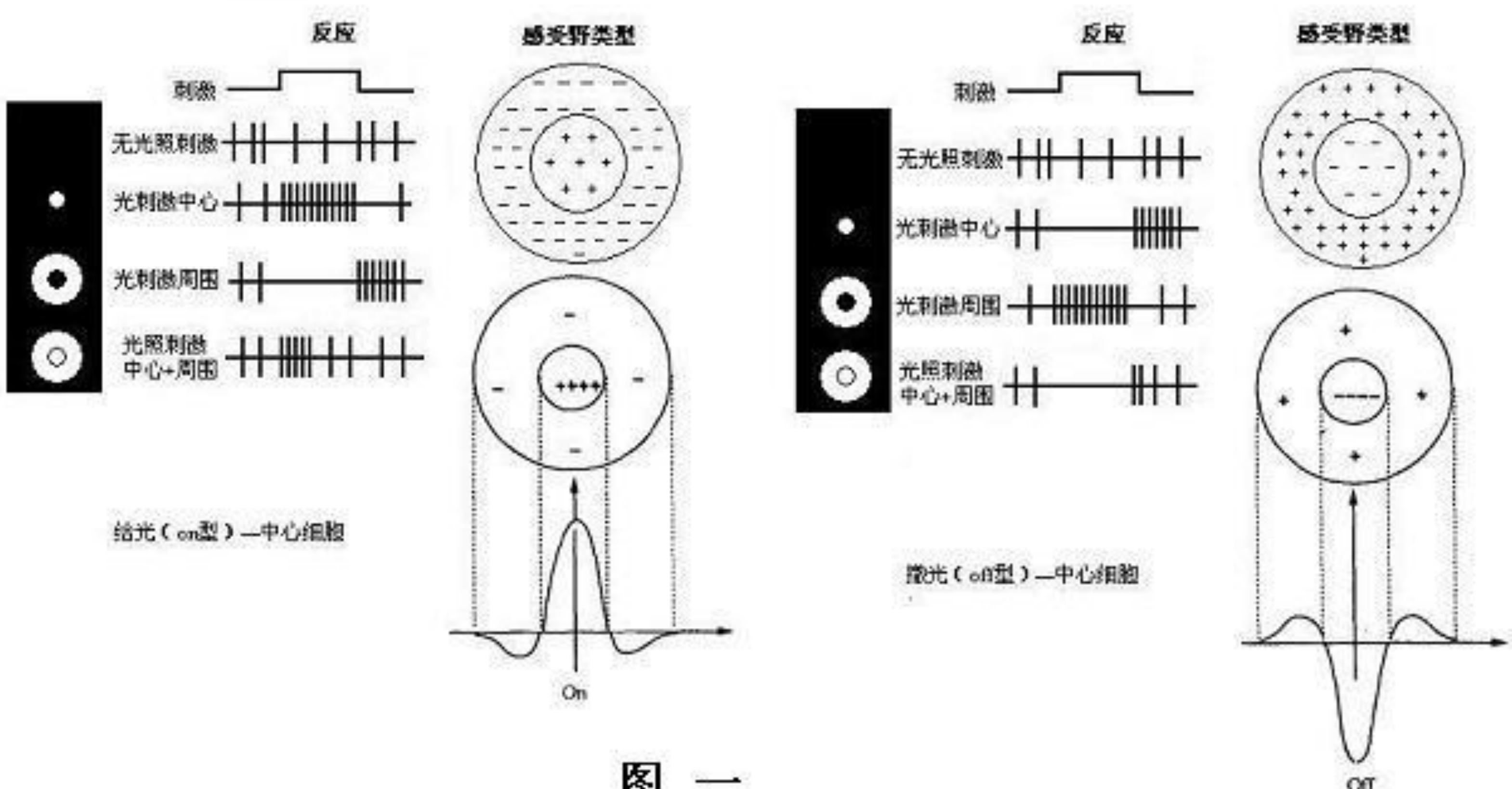
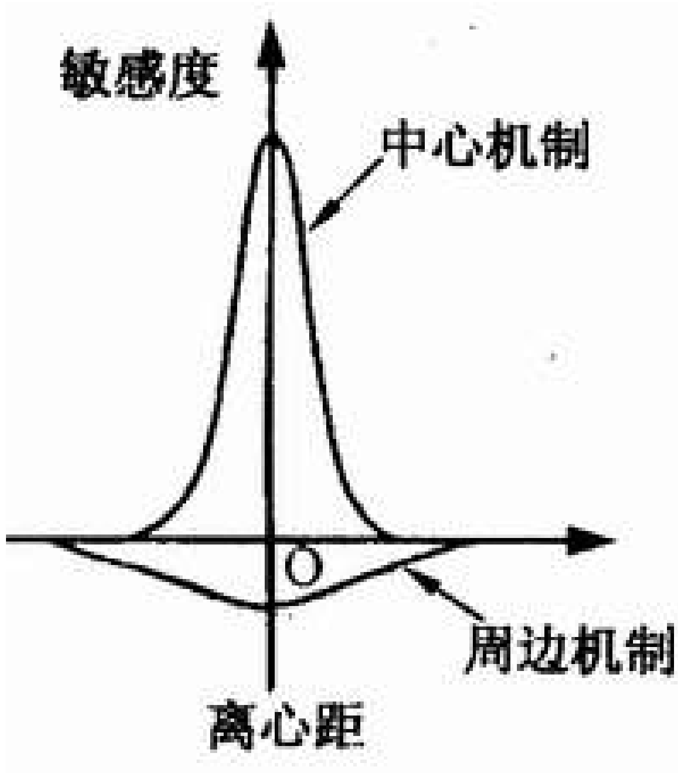
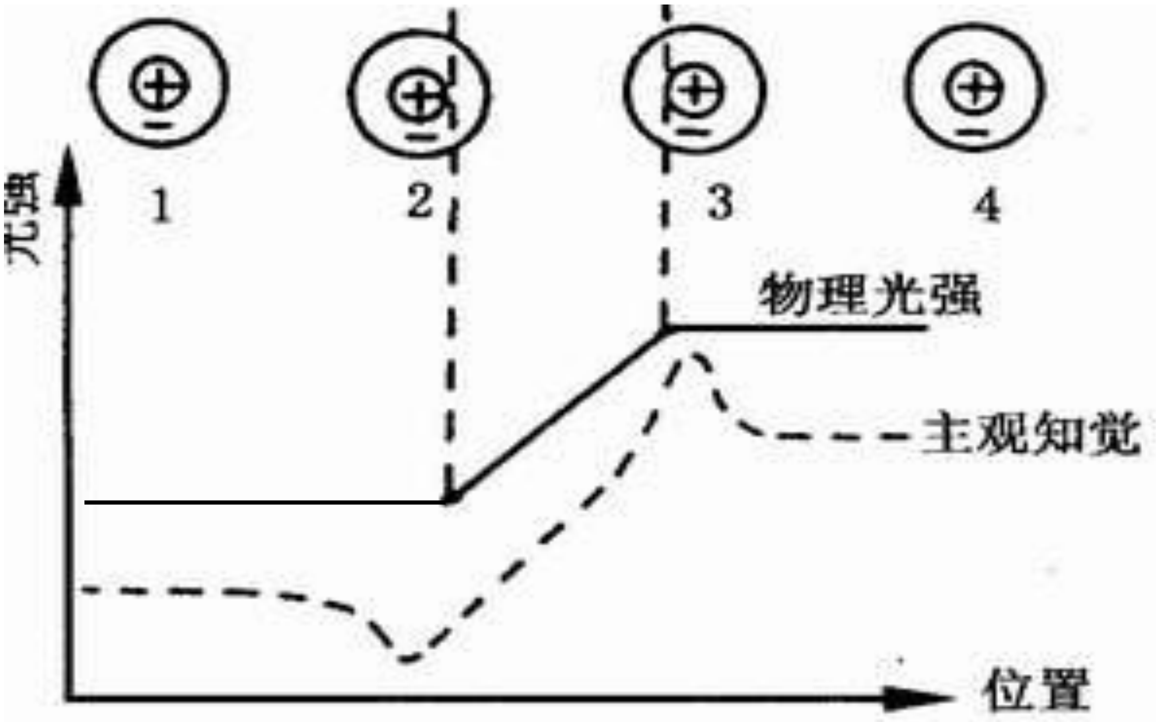

反应
感受野类型
刺激
无光照刺激
光刺激中心
光刺激周围
光照刺激
中心+周围
给光（on型）—中心细胞
On
反应
感受野类型
刺激
无光照刺激
光刺激中心
光刺激周围
光照刺激
中心+周围
撤光（off型）—中心细胞
Off

图 一

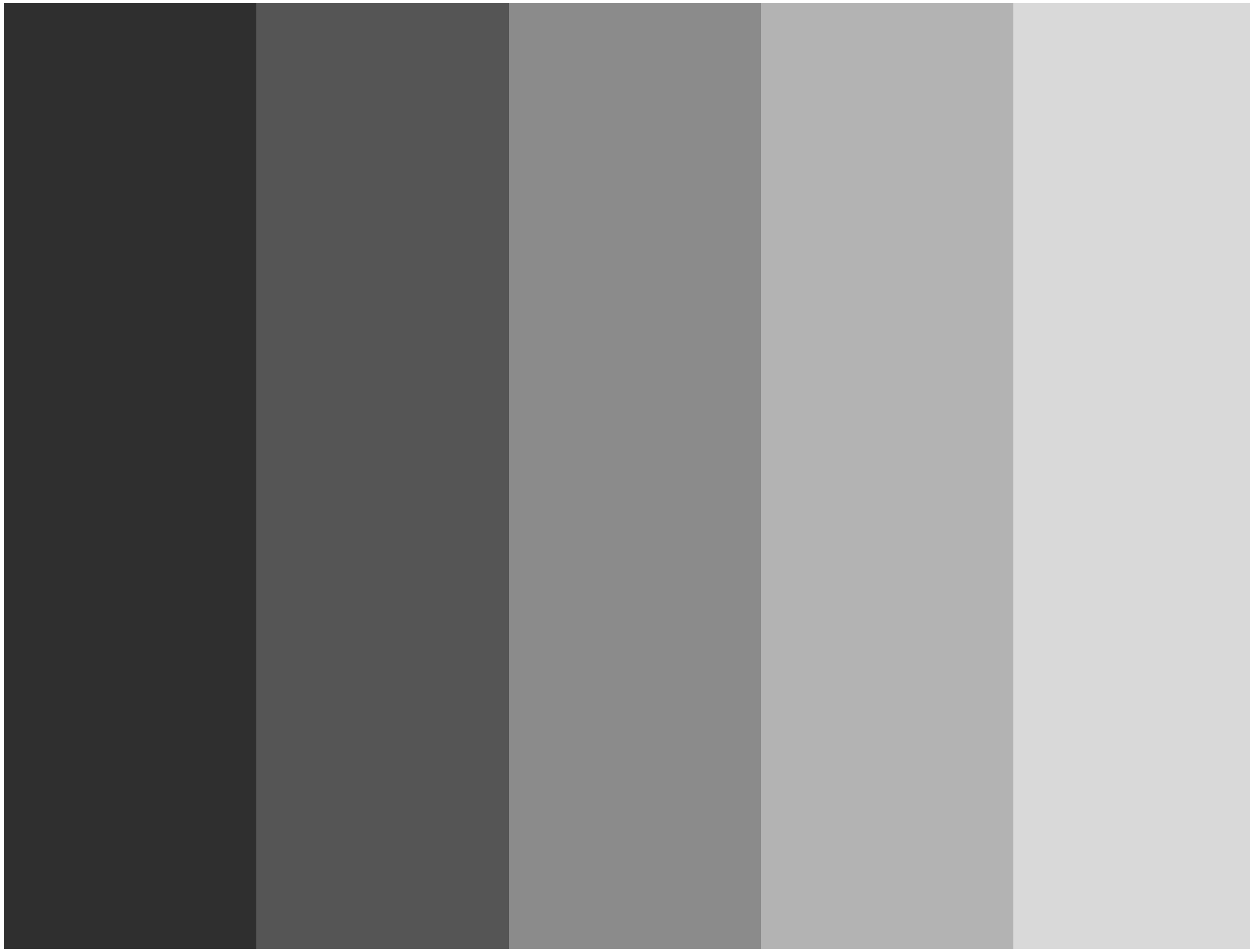
感受野同心圆拮抗式模型



（Rodieck, 1965）

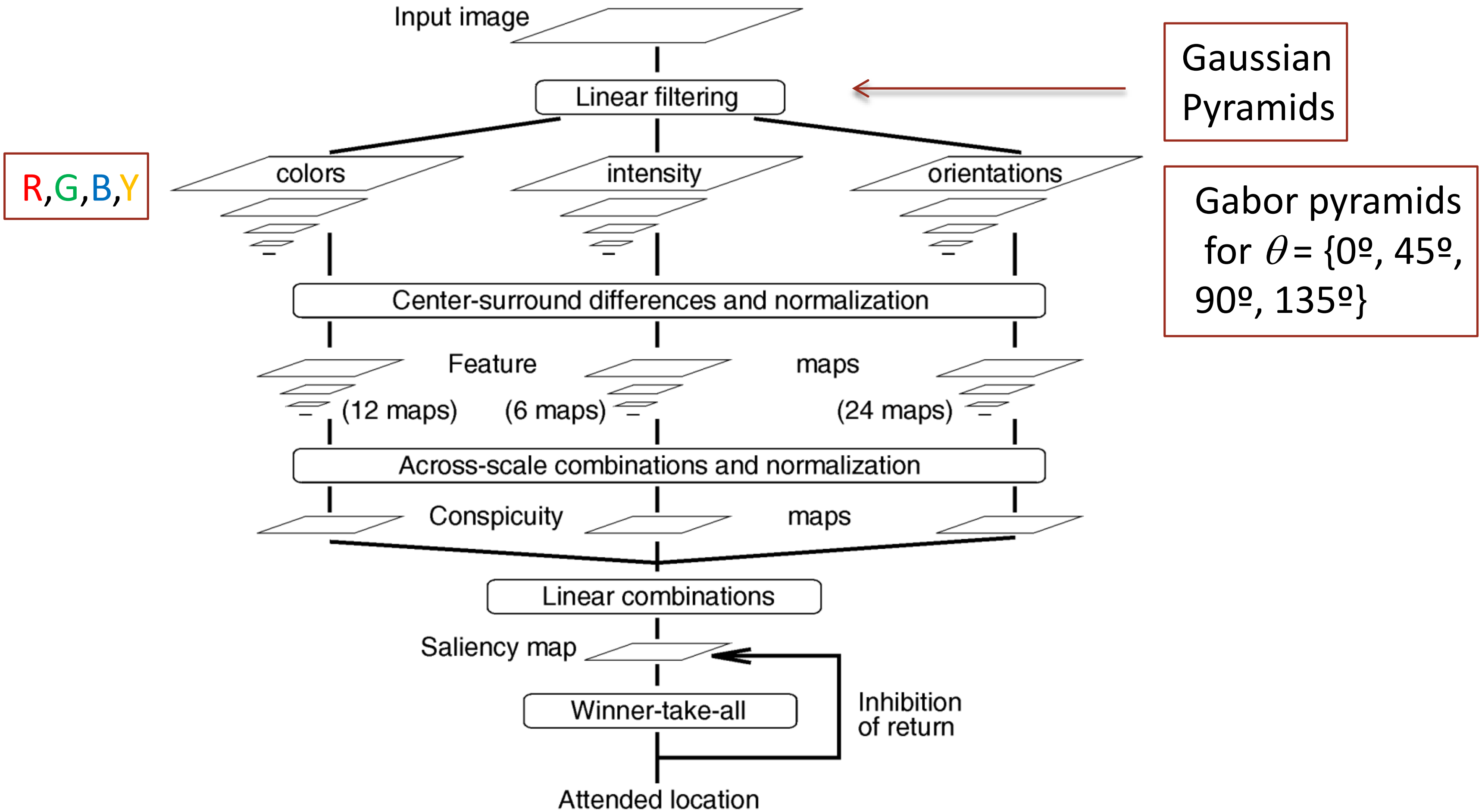


马赫带效应



马赫带一指人们在明暗变化的边界，常常在亮区看到一条更亮的光带，而在暗区看到一条更暗的线条。这就是马赫带现象，马赫带不是由于刺激能量的分布，而是由于受到视觉”惰性”的影响

L. Itti's approach (TPAMI,1998)



Laplacian of Gaussian (LOG)

- **The well-known Laplacian derivative operator (isotropic second derivative) is given by:**

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

- **For Gaussian pyramid, we have**

$$g(x;\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2/2\sigma^2}$$

- **Various derivatives for the Gaussian kernels can be derived and related to Laplacian pyramid**

Relations between DOG and LOG

- Various derivatives are:

$$\frac{dg(x;\sigma)}{dx} = \frac{-x}{\sigma^2} g(x;\sigma)$$

$$\frac{d^2 g(x;\sigma)}{d^2 x} = \left(\frac{x^2}{\sigma^2} - 1\right)\frac{1}{\sigma^2} g(x;\sigma)$$

$$\frac{dg(x;\sigma)}{d\sigma} = \left(\frac{x^2}{\sigma^2} - 1\right)\frac{1}{\sigma} g(x;\sigma)$$

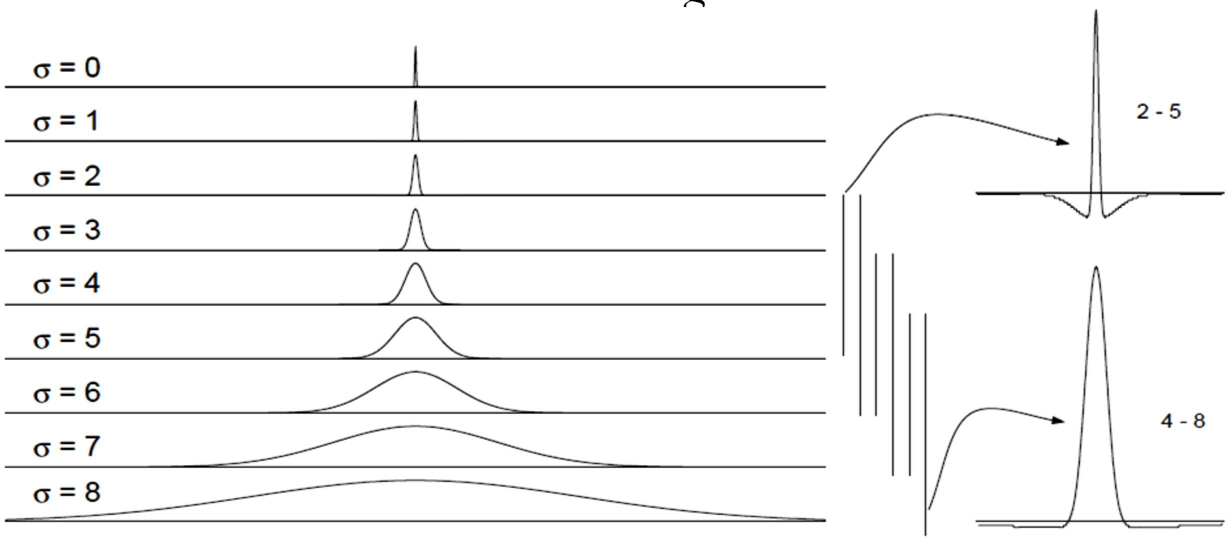
- Therefore,

$$\frac{d^2 g(x;\sigma)}{d^2 x} = c_0(\sigma)\frac{dg(x;\sigma)}{d\sigma}$$

$$\approx c_1(\sigma)[g(x;\sigma) - g(x;\sigma + \Delta\sigma)]$$

L. Itti's approach (TPAMI,1998)

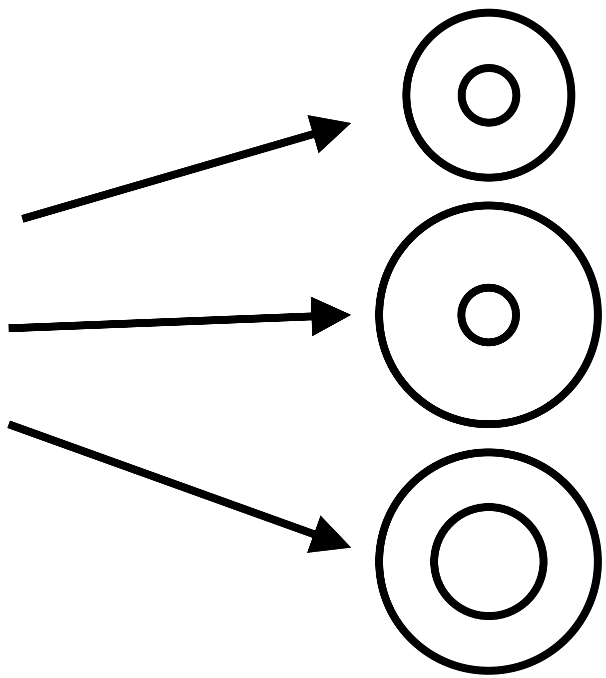
- **Center-surround Difference**
- Achieve center-surround difference through across-scale difference



- Operated denoted by Θ: Interpolation to finer scale and point-to-point subtraction
- One pyramid for each channel: $I(\sigma)$, $R(\sigma)$, $G(\sigma)$, $B(\sigma)$, $Y(\sigma)$ where $\sigma \in [0..8]$ is the scale

L. Itti's approach (TPAMI,1998)

- **Center-surround Difference**
  - **Intensity Feature Maps**
- $I(c, s) = |\ \mathrm{I}(c)\ \Theta\ \mathrm{I}(s)|$
- $c \in \{2, 3, 4\}$
- $s = c + \delta$ where $\delta \in \{3, 4\}$
- So $I(2, 5) = |\ \mathrm{I}(2)\ \Theta\ \mathrm{I}(5)|$
  $I(2, 6) = |\ \mathrm{I}(2)\ \Theta\ \mathrm{I}(6)|$
  $I(3, 6) = |\ \mathrm{I}(3)\ \Theta\ \mathrm{I}(6)|$
  …
- → 6 Feature Maps



L. Itti's approach (TPAMI,1998)

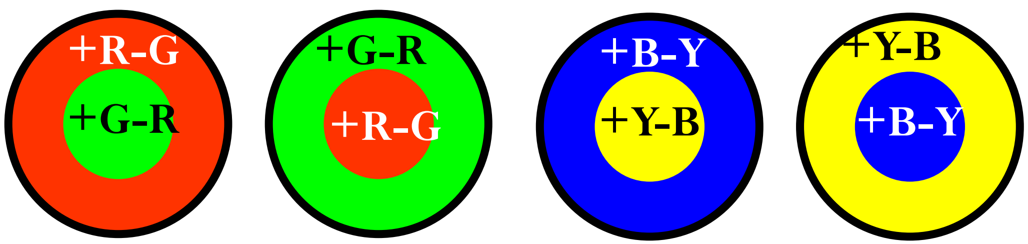
- Center-surround Difference
  - **Color Feature Maps**

Red-Green and Yellow-Blue

- Center-surround Difference
  - **Orientation Feature Maps**

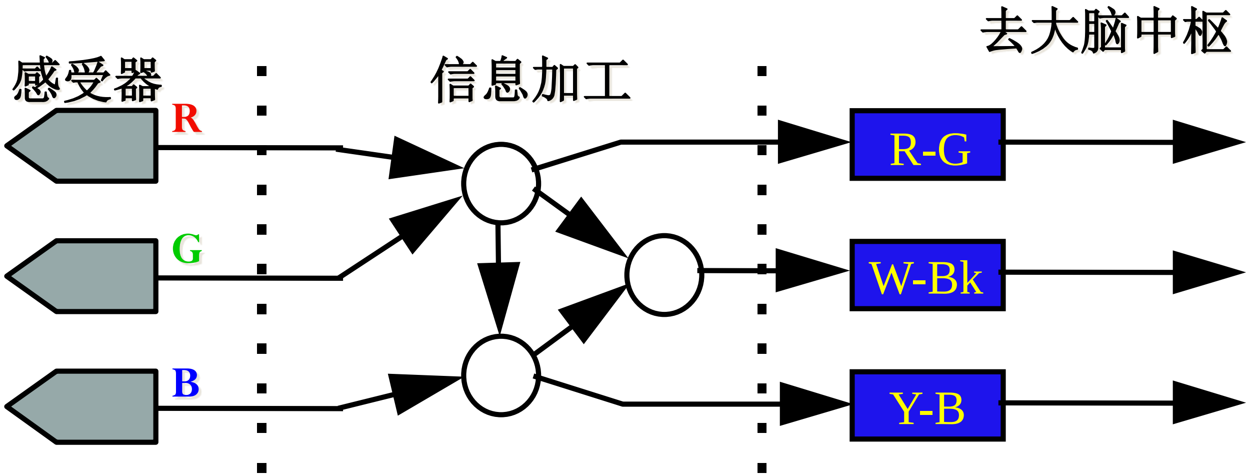
$$O(c,s,\theta) = \left|O(c,\theta) - O(s,\theta)\right|$$

Same c and s as with intensity



$$RG(c, s) = |\, (R(c) - G(c)) \,\Theta\, (G(s) - R(s)) \,|$$
$$BY(c, s) = |\, (B(c) - Y(c)) \,\Theta\, (Y(s) - B(s)) \,|$$

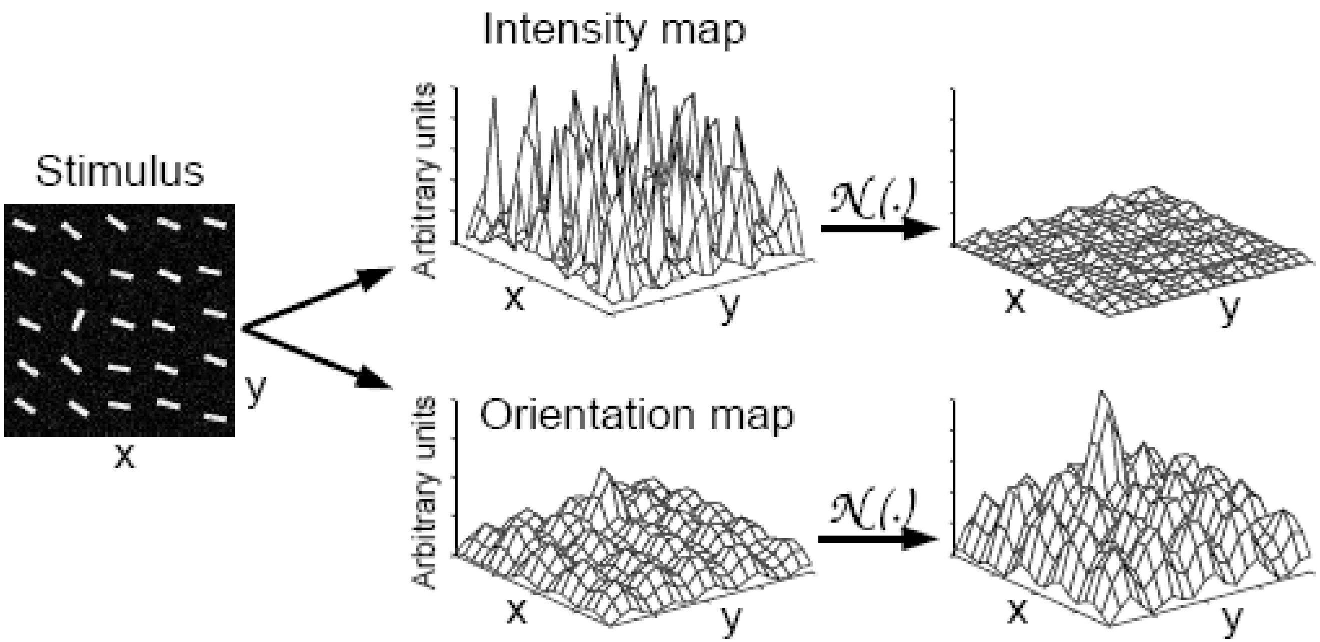
人的色觉-视觉通路



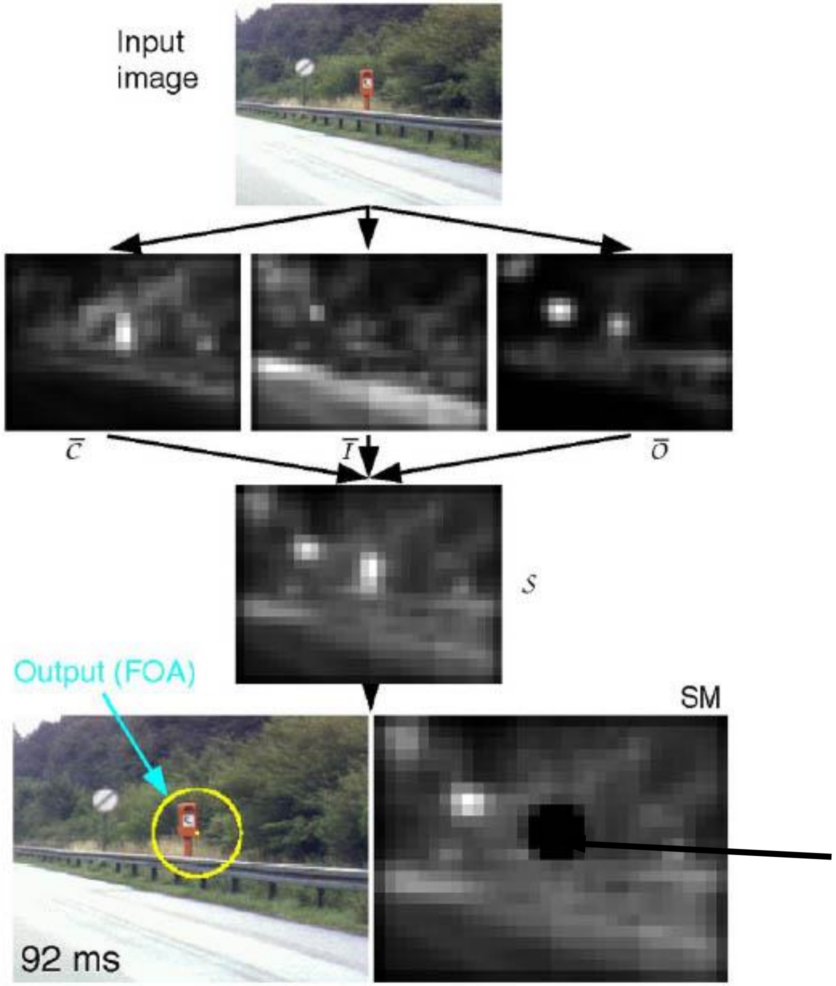
颜色视觉机制传输模型

L. Itti's approach (TPAMI,1998)

- **Normalization Operator**
  - Promotes maps with few strong peaks
  - Surpresses maps with many comparable peaks



L. Itti's approach (TPAMI,1998)



Inhibition of return

How to detect Saliency map?

- Button-up approach
  - L. Itti’s approach
  - Spectral Residual approach
  - Frequency-tuned approach
  - Global contrast based approach
- Top-down approach
  - Context-aware

Spectral Residual Approach(CVPR,2007)

- 从信息理论角度：信息可分为冗余部分和变化部分。人们的视觉对变化部分更敏感。视觉系统的一个基本原则就是抑制对频繁出现的特征的响应，同时对非常规的特征保持敏感。那么就将图像分为如下两部分：

$$\boxed{H(\text{Image}) = H(\text{Innovation}) + H(\text{Prior Knowledge}),}$$

新的信息，反映图像内容变化。

冗余信息，可以通过统计方法得到。

Spectral Residual Approach(CVPR,2007)

- 作者对图像的log频谱进行统计分析，得到了如下规律：大量图像的log频谱的平均值是和频率呈现正比关系的。

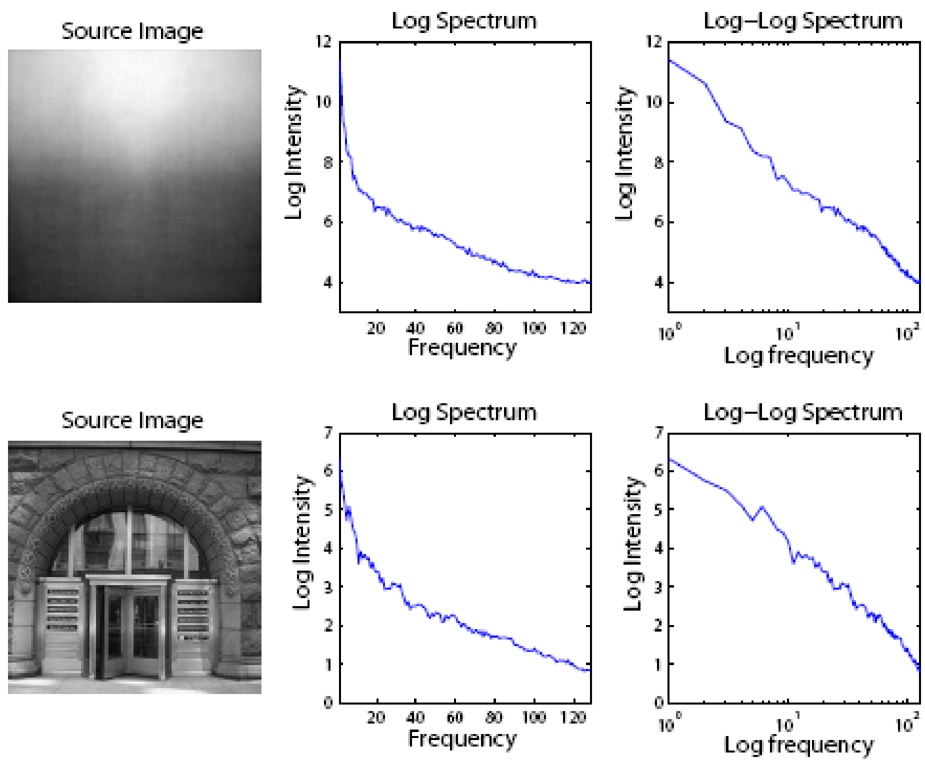


Figure 1. Examples of log spectrum and log-log spectrum. The first image is the average of 2277 natural images.

Spectral Residual Approach(CVPR,2007)

- 大量图像的log频谱和频率的曲线形状，在log-log scale上，几乎是一条直线。然后作者又提出了既然大量图像的log振幅谱都差不多趋近一条直线，那么一幅图像的log振幅谱减去平均log振幅谱不就是显著性部分了吗？这就是作者提出的：Spectral Residual理论。

$$\mathcal{R}(f) = \mathcal{L}(f) - \mathcal{A}(f).$$

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

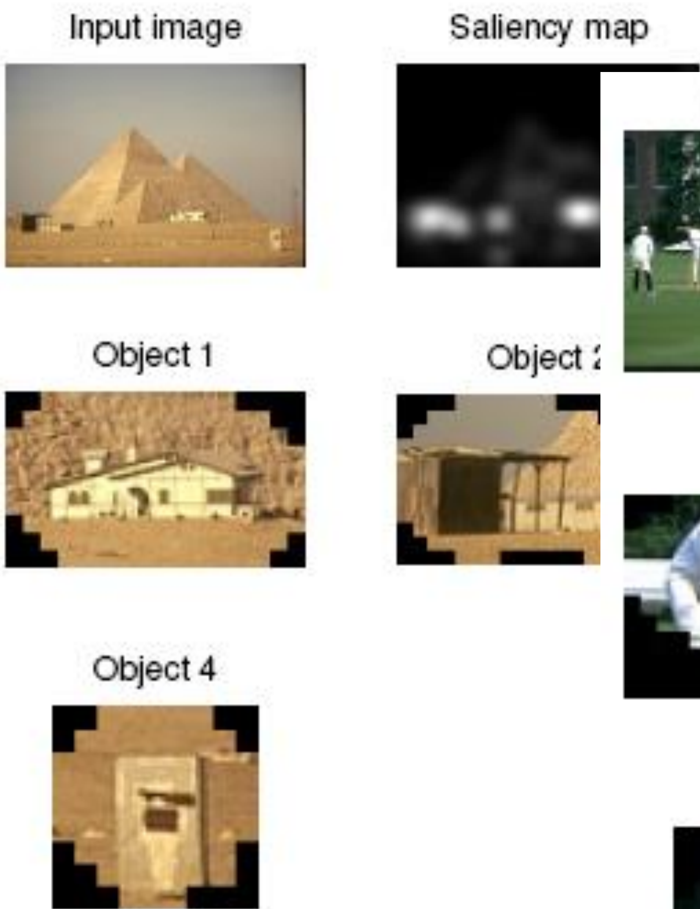
Spectral Residual Approach(CVPR,2007)

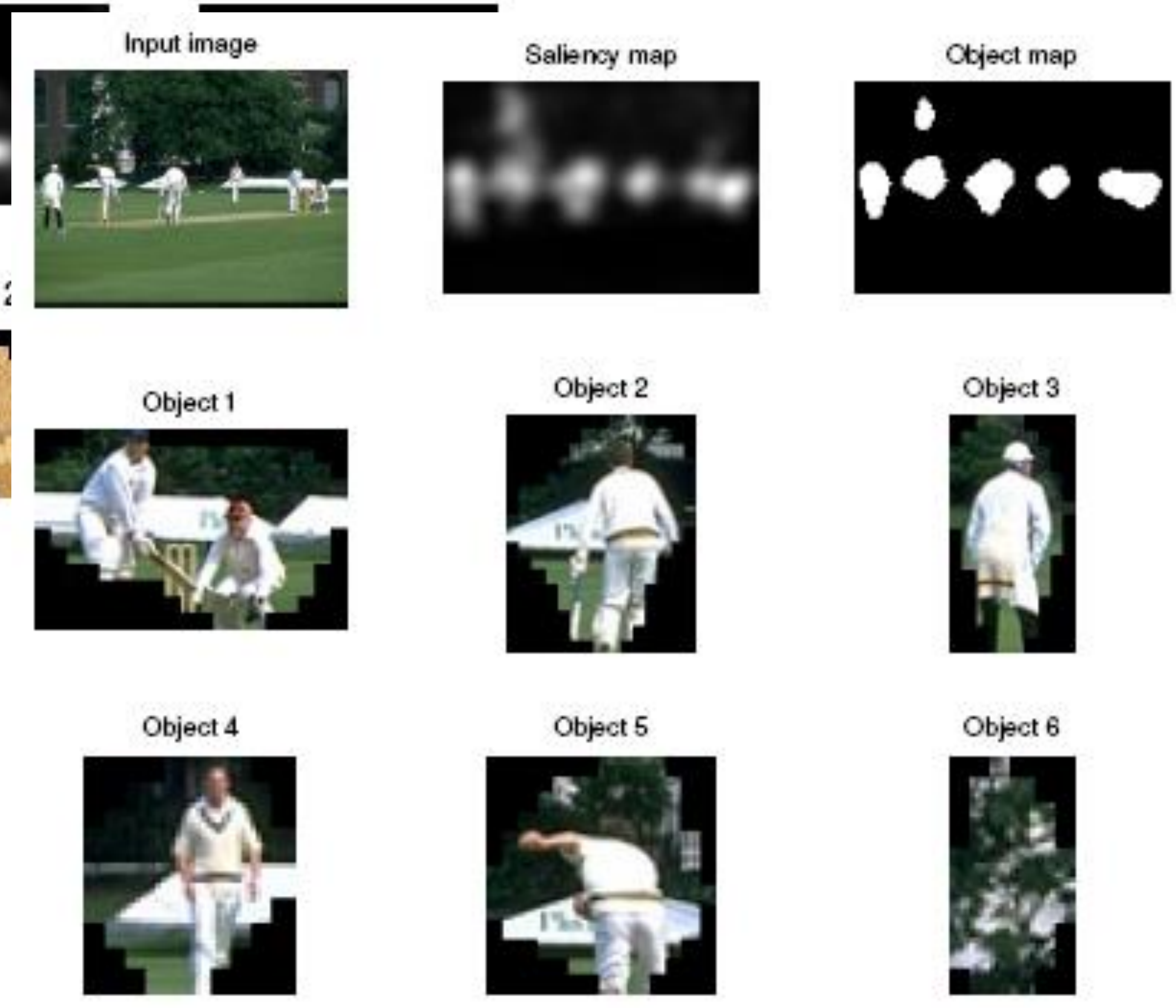
$$\mathcal{R}(f) = \mathcal{L}(f) - \mathcal{A}(f).$$

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

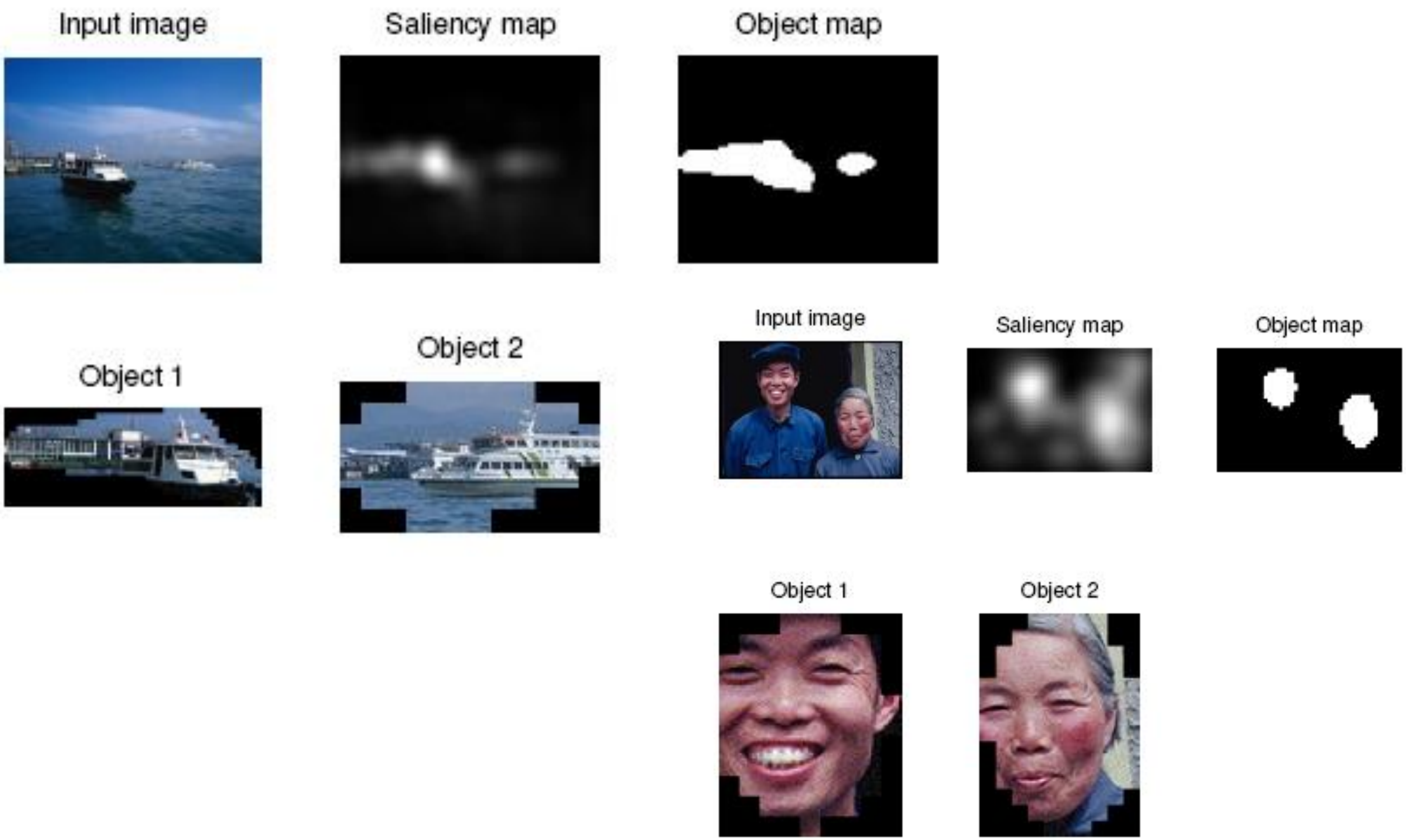
- First scaling image to 64x64.
- $h_n(f)$ is a local average filter (n=3 in this paper).
- Then the saliency map is smoothed with a gaussian filter g(x) (σ = 8).

Spectral Residual Approach(CVPR,2007)





Spectral Residual Approach(CVPR,2007)



Spectral Residual Approach(CVPR,2007)

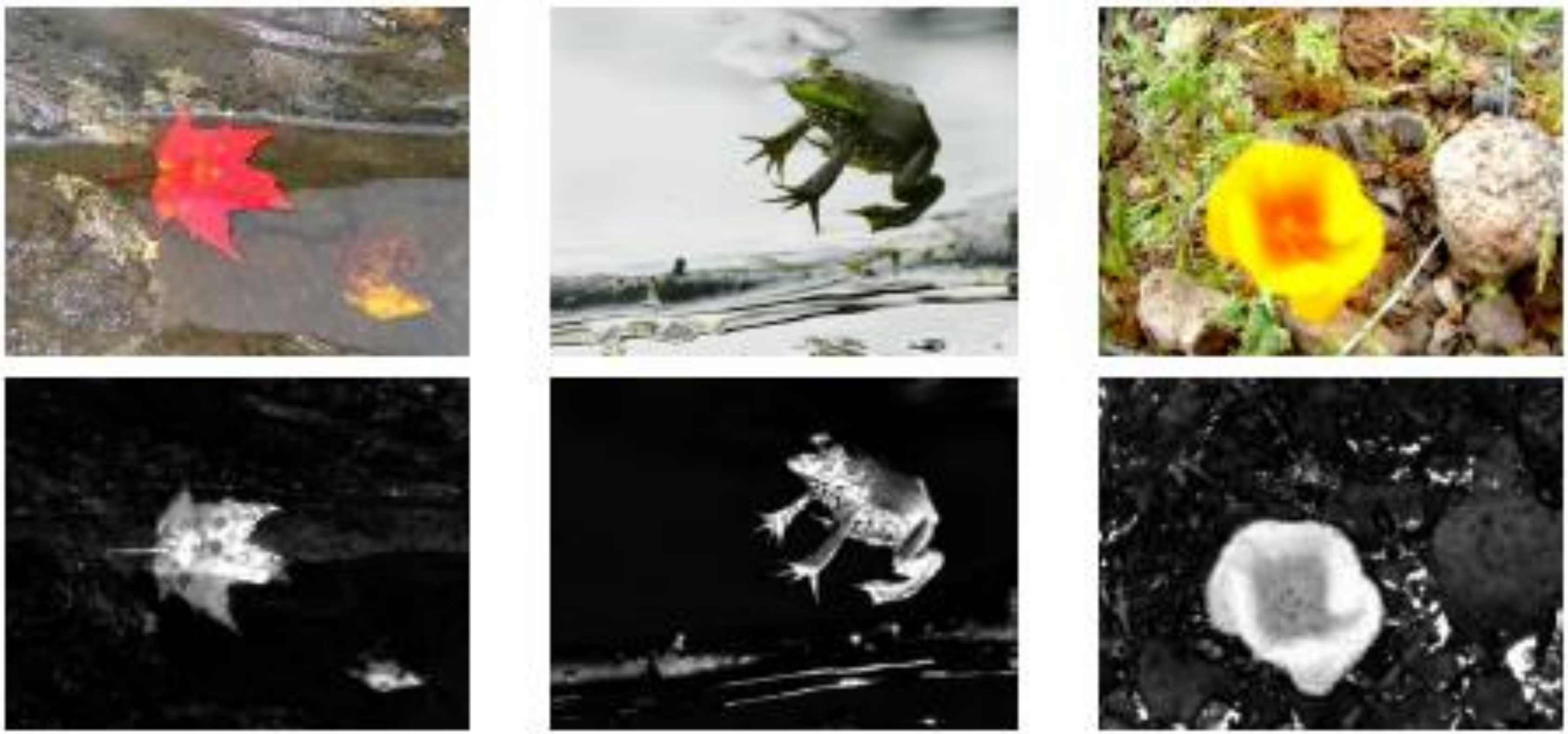
```
clear
Clc
%% Read image from file
inImg = im2double(rgb2gray(imread('256.png')));
%%inImg = imresize(inImg, 64/size(inImg, 2));
%% Spectral Residual
myFFT = fft2(inImg);
myLogAmplitude = log(abs(myFFT));
myPhase = angle(myFFT);
mySpectralResidual = myLogAmplitude - imfilter(myLogAmplitude, fspecial('average', 3),
    'replicate');
saliencyMap = abs(ifft2(exp(mySpectralResidual + i*myPhase))).^2;
%% After Effect
saliencyMap = mat2gray(imfilter(saliencyMap, fspecial('gaussian', [10, 10], 2.5)));
imshow(saliencyMap);
```

How to detect Saliency map?

- Button-up approach
  - L. Itti's approach
  - Spectral Residual approach
  - Frequency-tuned approach
  - Global contrast based approach

- Top-down approach
  - Context-aware

Frequency-tuned (CVPR,2009)



Frequency-tuned (CVPR,2009)

- Set the following requirements for a saliency detector:
  - Emphasize the largest salient objects.
  - Uniformly highlight whole salient regions.
  - Establish well-defined boundaries of salient objects.
  - Disregard high frequencies arising from texture, noise and blocking artifacts.
  - Efficiently output full resolution saliency maps.

Frequency-tuned (CVPR,2009)

- Choose the DoG filter for band pass filtering.

$$DoG(x,y) = \frac{1}{2\pi}\left[\frac{1}{\sigma_1^2}e^{-\frac{(x^2+y^2)}{2\sigma_1^2}} - \frac{1}{\sigma_2^2}e^{-\frac{(x^2+y^2)}{2\sigma_2^2}}\right]$$
$$= G(x,y,\sigma_1) - G(x,y,\sigma_2)$$

- Parameter selection
  - To implement a large ratio in standard deviations, σ_1 is driven to infinity.

Frequency-tuned (CVPR,2009)

Image Average

$$I_{\mu} = \begin{bmatrix} L_{\mu} \\ a_{\mu} \\ b_{\mu} \end{bmatrix}$$

$$S(x, y) = \left\| I_{\mu} - I_{\omega_{hc}}(x, y) \right\|$$

Gaussian blur

$$I_{\omega_{hc}}(x, y) = \begin{bmatrix} L_{\omega_{hc}} \\ a_{\omega_{hc}} \\ b_{\omega_{hc}} \end{bmatrix}$$

Frequency-tuned (CVPR,2009)

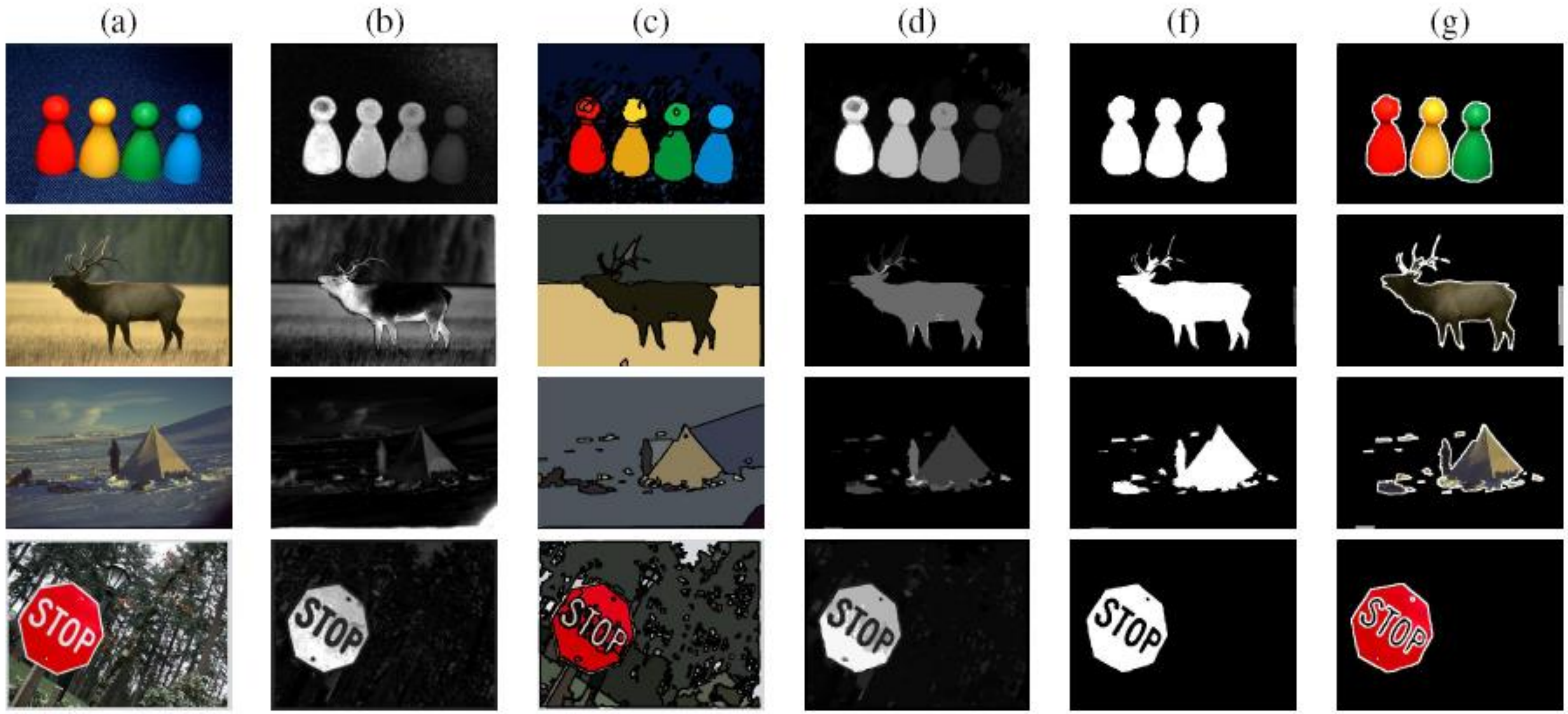


Figure 6. (a) is the original image. The average saliency per segment (d) is computed using the saliency map (b) and the mean-shift segmented image (c). Those segments that have a saliency value greater than the adaptive threshold computed in Eq. 9 are assigned ones (white) and the rest zeroes (black) in (f). The salient objects resulting from binary map (f) are shown in (g).

Frequency-tuned (CVPR,2009)

```
% Read image and blur it with a 3x3 or 5x5 Gaussian filter
img = imread('input_image.jpg');%Provide input image path
gfrgb = imfilter(img, fspecial('gaussian', 3, 3), 'symmetric', 'conv');
% Perform sRGB to CIE Lab color space conversion (using D65)
cform = makecform('srgb2lab', 'whitepoint', whitepoint('d65'));
lab = applycform(gfrgb,cform);
% Compute Lab average values (note that in the paper this
% average is found from the unblurred original image, but the results are quite similar)
l = double(lab(:,:,1)); lm = mean(mean(l));
a = double(lab(:,:,2)); am = mean(mean(a));
b = double(lab(:,:,3)); bm = mean(mean(b));
% Finally compute the saliency map and display it.
sm = (l-lm).^2 + (a-am).^2 + (b-bm).^2;
imshow(sm,[]);
```

How to detect Saliency map?

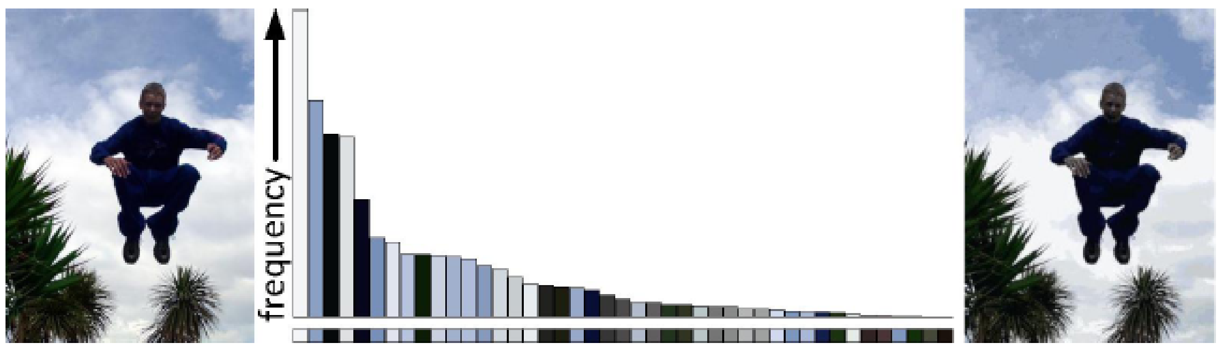
- Button-up approach
  - L. Itti's approach
  - Spectral Residual approach
  - Frequency-tuned approach
  - Global contrast based approach

- Top-down approach
  - Context-aware

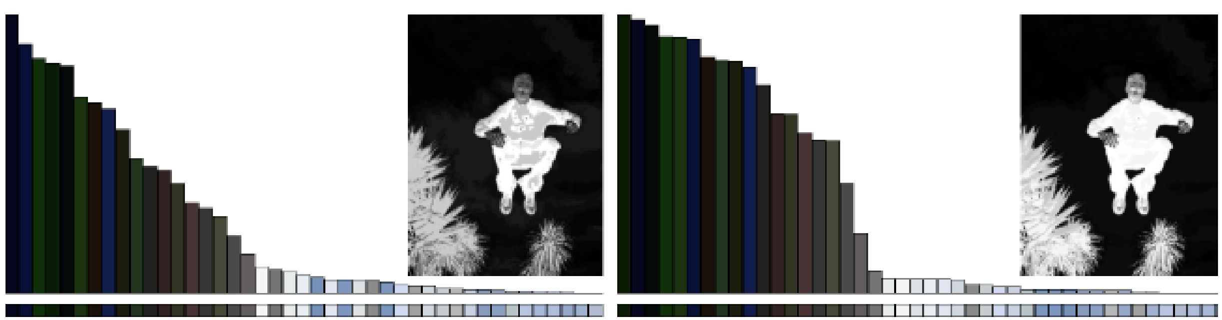
Global contrast-based (CVPR,2011)

Histogram Based Contrast (HC)

$$S(c_l) = \sum_{j=1}^{n} f_j D(c_l, c_j)$$



Histogram based speed up



Color space smoothing

Global contrast-based (CVPR,2011)

Region Based Contrast (RC)

Segmentation $\qquad \sigma_s^2 \to \infty \qquad \sigma_s^2 = 0.4$

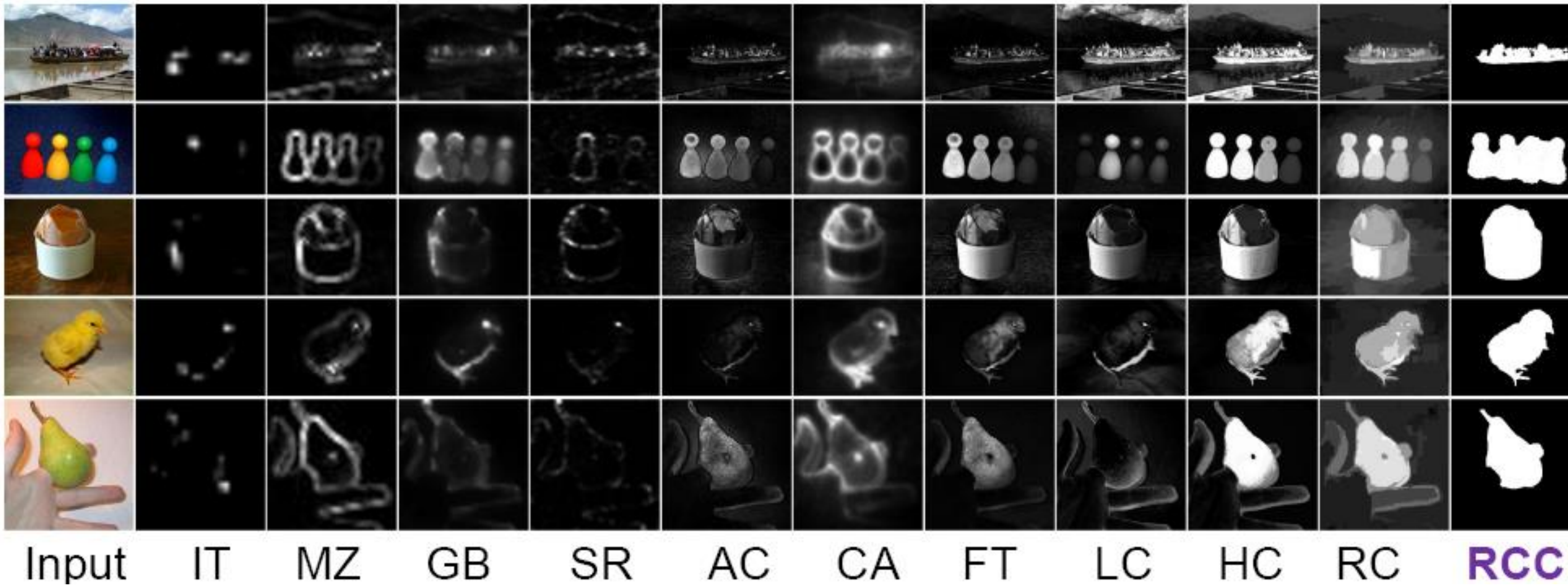
Spatial weighting

Region size

$$S(r_k) = \sum_{r_k \neq r_i} \exp\left(-\frac{D_s(r_k, r_i)}{\sigma_s^2}\right) \omega(r_i) D_r(r_k, r_i)$$

Region contrast by sparse histogram comparison.

Global contrast-based (CVPR,2011)



- **RC based saliency Cut achieves P = 90%, R = 90%, compared to previous best results P = 75%, R = 83% on this dataset.**
- **P = Precision; R = Recall.**

How to detect Saliency map?

- Button-up approach
  - L. Itti’s approach
  - Spectral Residual approach
  - Frequency-tuned approach
  - Global contrast based approach

- Top-down approach
  - Context-aware

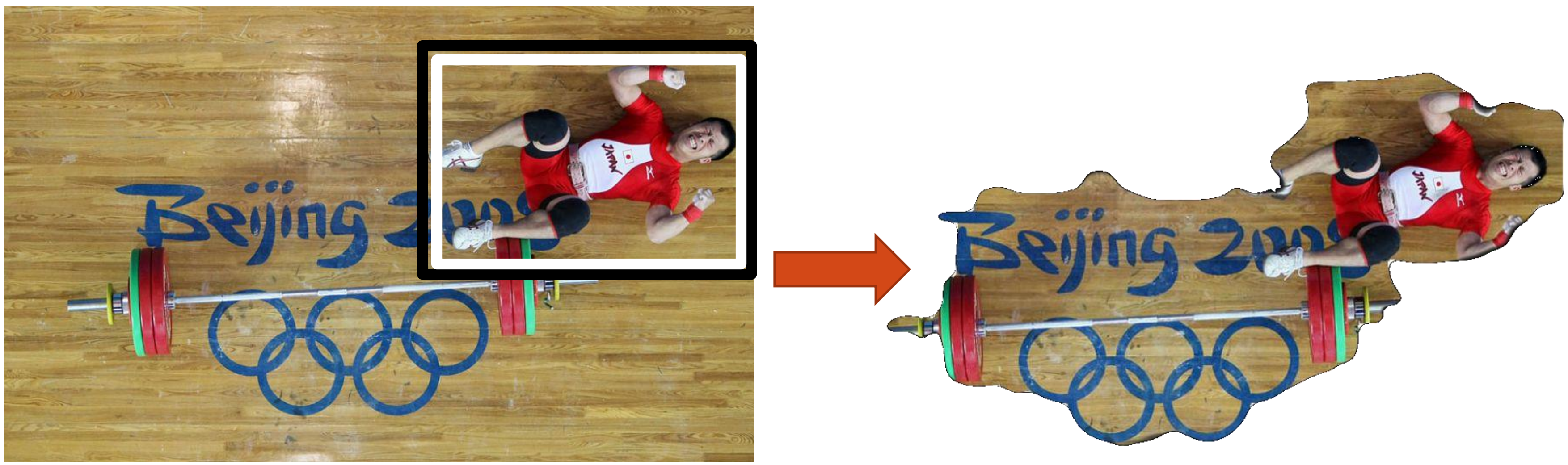
How to detect Saliency map?

- Button-up approach
  - L. Itti's approach
  - Spectral Residual approach
  - Frequency-tuned approach
  - Global contrast based approach

- Top-down approach
  - Context-aware

Context-Aware (CVPR,2010)

- **Goal:** Convey the image content



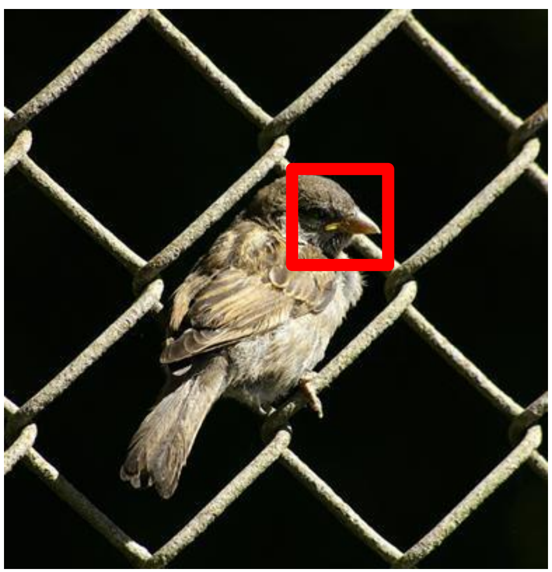
Context-Aware (CVPR,2010)

- Principles of context-aware saliency:
  - 1. Local low-level considerations, including factors such as contrast and color.
  - 2. Global considerations, which suppress frequently occurring features, while maintaining features that deviate from the norm.
  - 3. Visual organization rules, which state that visual forms may possess one or several centers of gravity about which the form is organized.
  - 4. High-level factors, such as human faces.

Context-Aware (CVPR,2010)

- Distance between a pair of patches:

$$d(p_i, p_j) = \frac{d_{color}(p_i, p_j)}{1 + c \cdot d_{position}(p_i, p_j)}$$



High $\forall j$

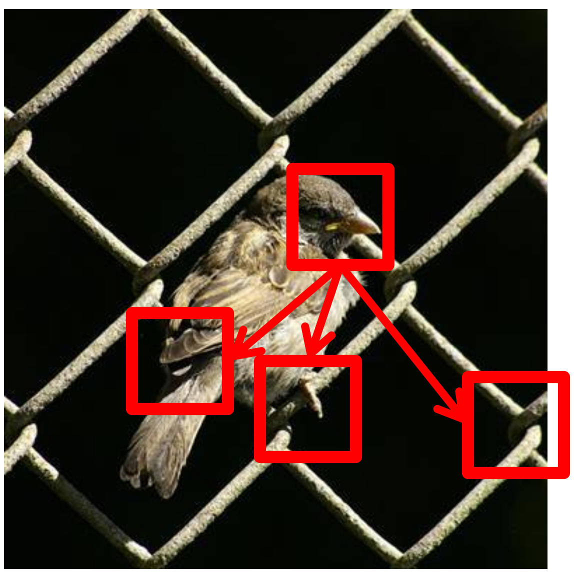
salient

Context-Aware (CVPR,2010)

- Distance between a pair of patches:

$$S_i^r = 1 - \exp\left[-\frac{1}{K} \sum_{k=1}^{K} d(p_i^r, q_j^r) \right]$$

$q_k^r =$ K most similar patches at scale r



High for K most similar

$\updownarrow$

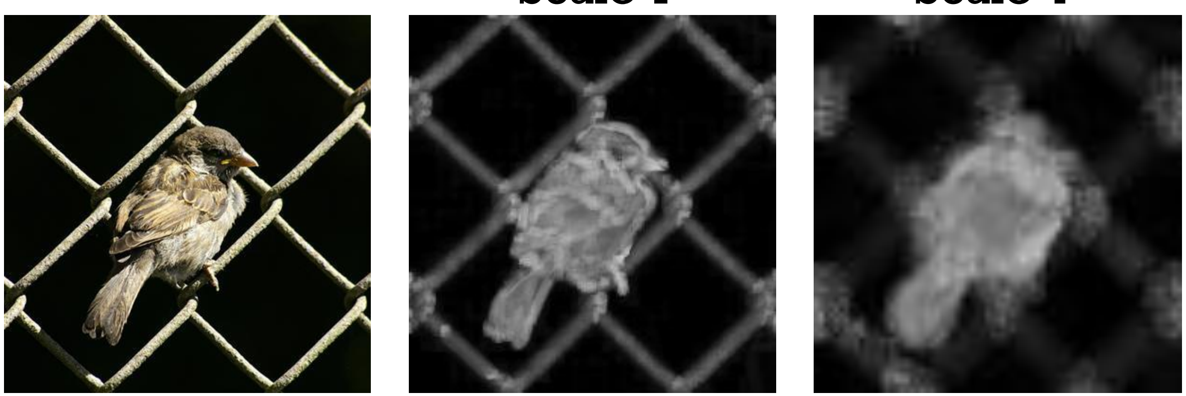
Saliency

Context-Aware (CVPR,2010)

- Salient at:
  - Multiple scales ➔ foreground
  - Few scales ➔ background

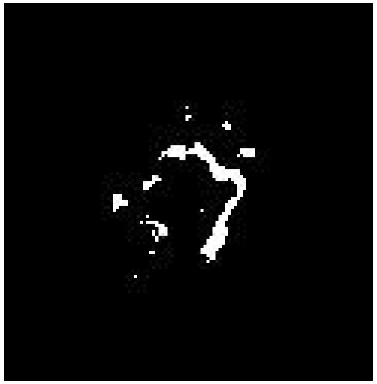
$$\bar{S}_i = \frac{1}{M}\sum_{r=r_1}^{r_M} S_i^r$$

Scale 1 Scale 4



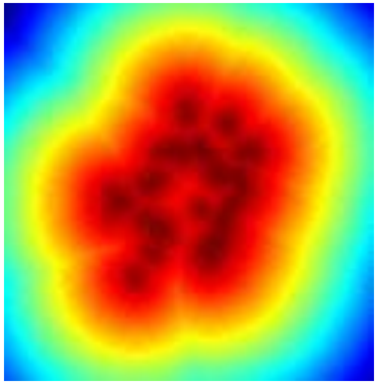
Context-Aware (CVPR,2010)

- Foci = $\bar{S}_i > 0.8$

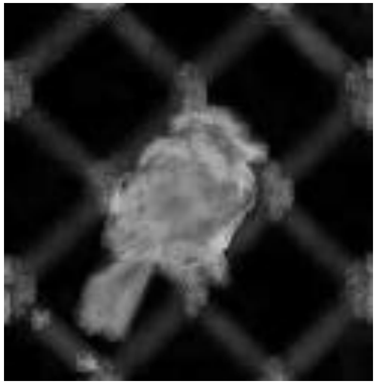


- Include distance map

$1 - d_{foci}(i)$ $\qquad$ $\bar{S}_i$



X

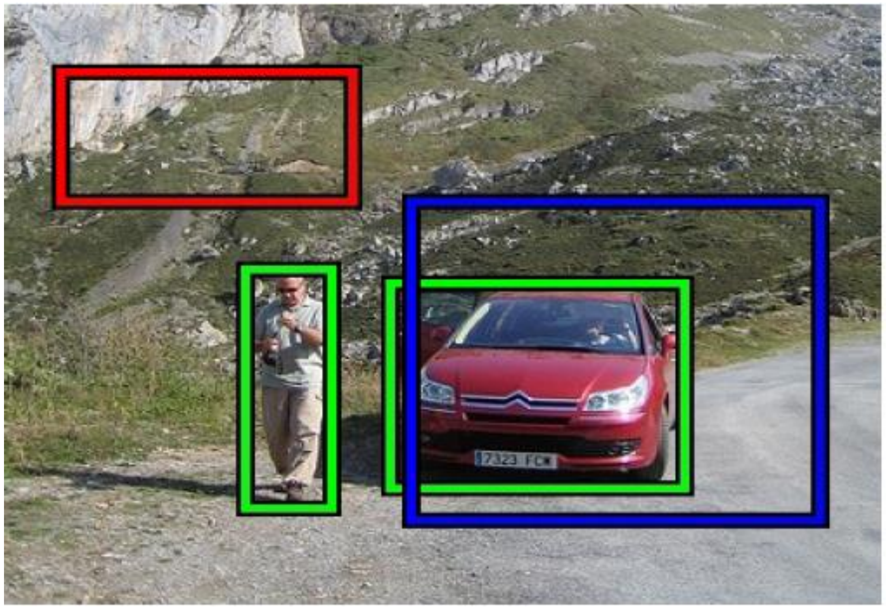


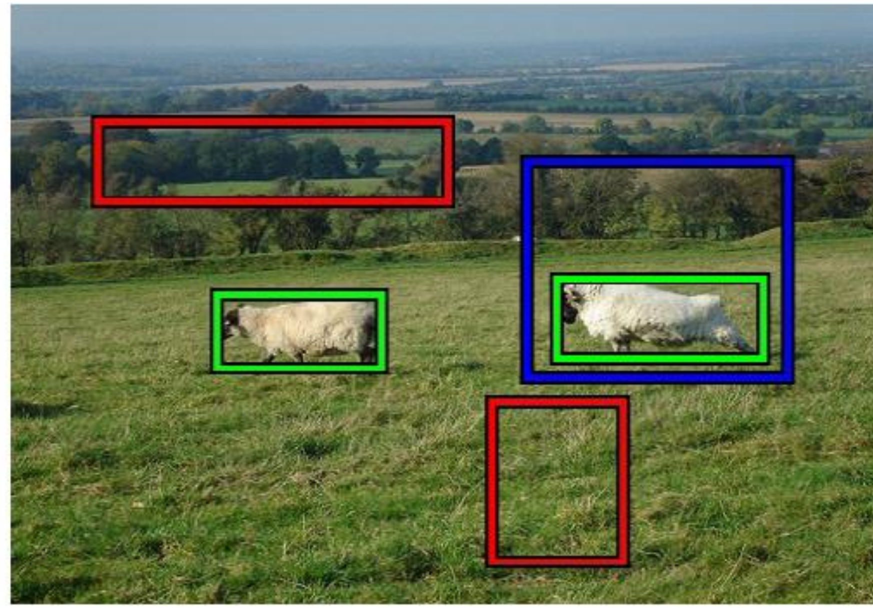
$$\hat{S}_i = \bar{S}_i \left(1 - d_{foci}(i)\right)$$

Context-Aware (CVPR,2010)

- High-level factors:
  - The saliency map should be further enhanced using some high-level factors, such as recognized objects or face detection.
  - In this paper, the face detection algorithm is incorporated , which generates 1 for face pixels and 0 otherwise.
  - The saliency map is modified by taking the maximum value of the saliency map and the face map.

似物性采样？





对于一幅480*640的图像，要识别其中的目标：

- 滑动窗口搜索：需搜索10^5−10^6个窗口
- 目前的似物性采样算法：需10^3−10^4个窗口。

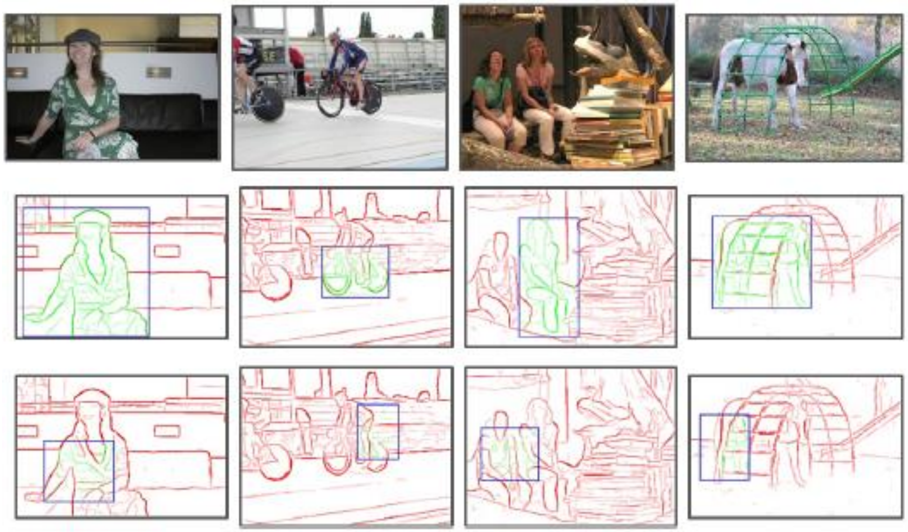
似物性采样研究现状及方法

表1 目前流行的似物性采样算法性能对比

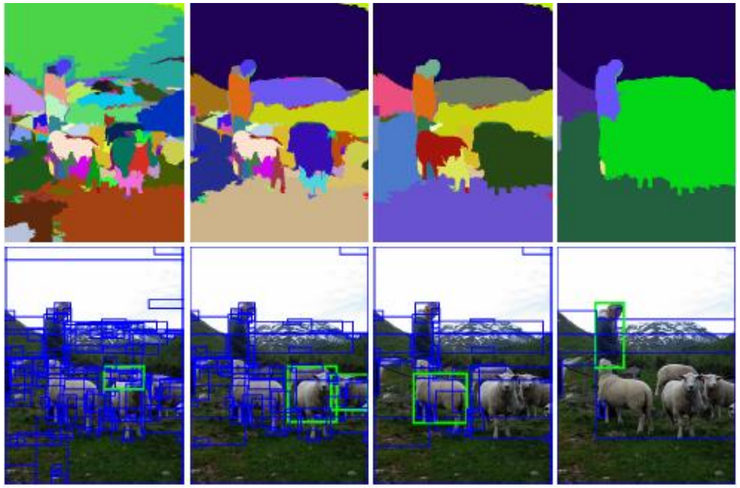
算法	发表期刊或会议	类型	是否输出分割	是否输出得分	计算时间(s)	召回率
Bing	CVPR2014	窗口评分	否	是	0.2	一般
CPMC	CVPR2010	分组合并	是	是	250	良好
EdgeBoxes	ECCV2014	窗口评分	否	是	0.3	优秀
Endres	ECCV2010	分组合并	是	否	100	优秀
Geodesic	ECCV2014	分组合并	是	是	1	优秀
MCG	T-PAMI2016	分组合并	是	是	30	优秀
Objectness	T-PAMI2012	窗口评分	否	是	3	一般
Rahtu	ICCV2011	窗口评分	否	否	3	-
RandomizedPrim's	ICCV2013	分组合并	是	否	1	一般
Rantalankila	CVPR2014	分组合并	是	否	10	-
Rigor	CVPR2014	分组合并	是	否	10	良好
SelectiveSearch	IJCV2013	分组合并	是	是	10	优秀

似物性采样研究现状及方法

目前基于RGB图像的似物性采样研究

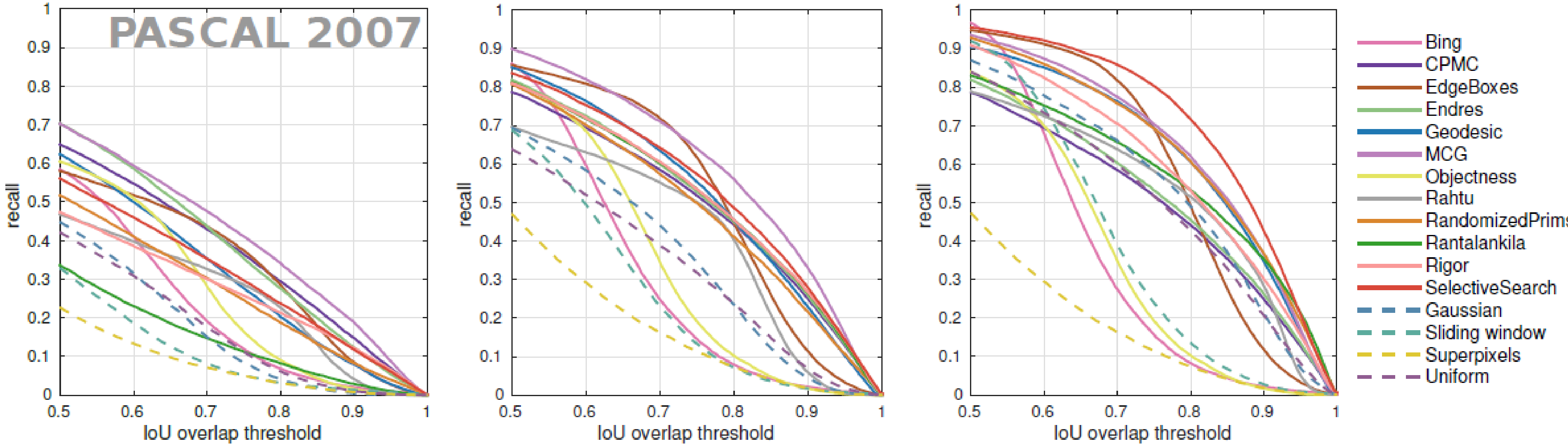


基于窗口评分的方法

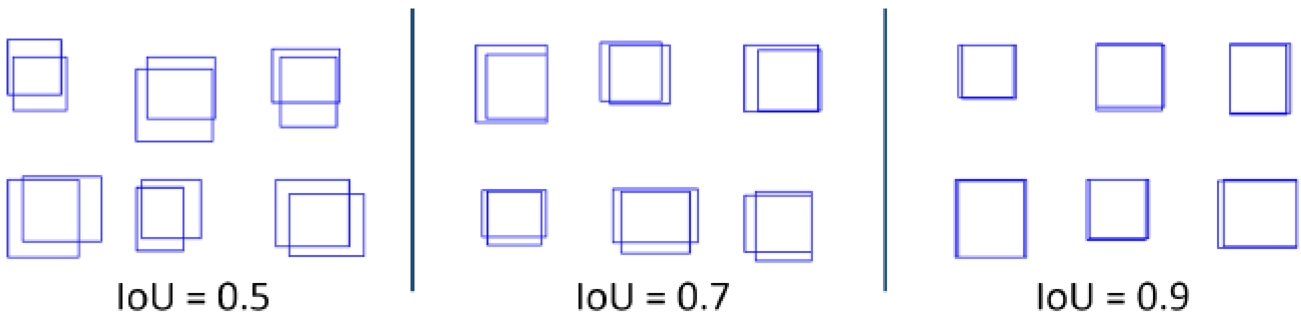


基于分组合并的方法

似物性采样VS滑动窗口搜索



(a) 100 proposals per image. (b) 1 000 proposals per image. (c) 10 000 proposals per image.



似物性采样在目标识别中的作用

	aero	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
Bing	56.6	54.9	45.0	28.6	24.6	53.9	63.5	72.5	15.6	59.4	49.0	59.7	68.5	60.3	50.7	16.5	49.0	42.8	64.8	44.9	49.0
CPMC	65.2	61.8	58.2	37.2	17.9	71.0	67.3	76.7	22.9	61.2	64.6	70.1	77.0	69.2	54.8	18.5	52.6	63.4	71.7	61.5	57.1
EdgeBoxes	67.0	69.9	59.8	46.1	28.3	72.9	72.3	73.8	28.8	68.1	62.4	67.6	79.2	73.6	62.4	28.2	55.8	61.2	70.4	59.7	60.4
Endres	61.5	70.8	57.1	33.5	18.0	72.5	68.8	77.3	21.7	61.8	64.5	68.2	78.0	69.9	56.2	21.4	54.5	63.2	72.4	56.9	57.4
Geodesic	63.2	68.0	55.9	39.2	19.8	71.1	70.4	74.4	24.8	65.0	63.5	65.6	78.7	69.2	58.0	20.4	54.5	57.8	70.2	60.9	57.5
MCG	66.6	69.1	60.1	42.0	28.5	71.9	72.3	77.3	30.2	61.3	62.4	69.8	77.4	68.2	62.2	27.5	57.6	66.0	75.8	59.4	60.3
Objectness	62.4	61.5	51.0	32.0	19.3	65.8	64.3	69.5	18.0	55.4	51.4	60.1	74.1	64.7	50.9	17.3	41.9	50.9	67.8	49.0	51.4
Rahtu	62.8	60.9	53.3	35.1	15.3	72.6	60.5	75.1	15.4	56.9	61.6	66.3	76.3	65.2	51.2	14.1	44.6	58.1	72.0	54.3	53.6
RandomizedPrims	70.2	68.2	55.5	39.5	18.5	72.3	63.7	76.8	25.7	62.4	64.2	68.7	76.6	68.5	51.0	22.4	53.1	62.9	72.4	59.7	57.6
Rantalankila	64.7	66.1	57.2	37.8	19.7	74.2	67.5	78.2	23.0	63.6	63.4	70.3	78.6	69.8	55.9	21.4	50.8	64.3	74.1	58.3	57.9
Rigor	62.6	70.5	57.5	40.1	15.9	72.9	65.7	77.9	28.6	65.1	63.7	68.6	77.9	68.9	54.8	23.3	56.3	63.8	73.7	60.3	58.4
SelectiveSearch	70.3	66.9	61.5	42.2	21.7	68.3	68.7	76.3	27.5	65.9	67.0	69.8	75.5	68.9	57.9	24.6	53.6	63.7	76.0	62.4	59.5
Gaussian	53.9	66.1	46.6	24.6	10.0	66.6	52.2	77.1	20.6	48.7	64.1	65.5	75.6	64.2	47.0	14.2	38.1	58.2	70.5	53.0	50.8
SlidingWindow	42.0	57.7	40.1	23.7	9.3	60.8	47.8	72.8	12.5	42.1	44.7	63.7	72.8	62.5	44.5	8.5	34.3	47.7	62.3	46.6	44.8
Superpixels	29.7	5.5	19.8	10.4	9.0	7.4	24.4	42.0	15.1	39.9	6.6	30.3	10.7	13.7	12.8	8.9	40.7	18.1	4.9	55.6	20.3
Uniform	51.0	58.0	38.6	24.6	11.7	64.3	50.9	72.3	14.8	43.4	62.6	63.4	73.9	59.3	43.4	10.8	27.5	60.4	69.0	38.3	46.9
best per class	70.3	70.8	61.5	46.1	28.5	74.2	72.3	78.2	30.2	68.1	67.0	70.3	79.2	73.6	62.4	28.2	57.6	66.0	76.0	62.4	62.1

mean
49.0
57.1
60.4
57.4
57.5
60.3
51.4
53.6
57.6
57.9
58.4
59.5
50.8
44.8
20.3
46.9
62.1

R-CNN分类器在PASCAL VOC 2007目标识别数据集中，应用不同似物性采样算法下的识别精度。