Top-down Visual Attention Computational Model Using Visual Feature Distribution of Search Target

TOSHIYA OHIRA, TAKATSUGU HIRAYAMA,
SHOHEI USUI, SHOTA SATO,
KENJI MASE

Graduate School of Information Science,
Nagoya University

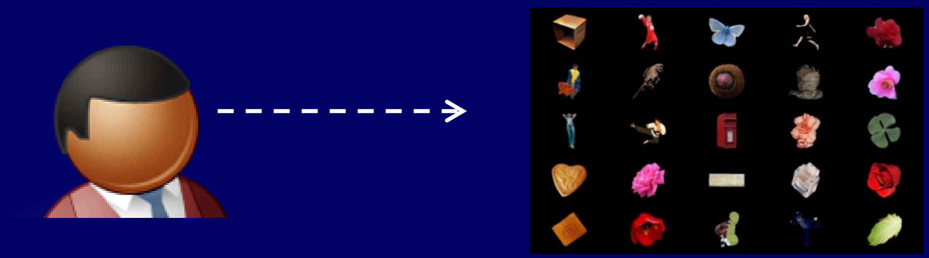
Backgrounds

- Gaze based Human Computer Interaction
  - A human-friendly robot
    - Establish joint attention and mutual gaze with humans
  - Driving support system
    - Estimate the visibility of signboards and guide plates

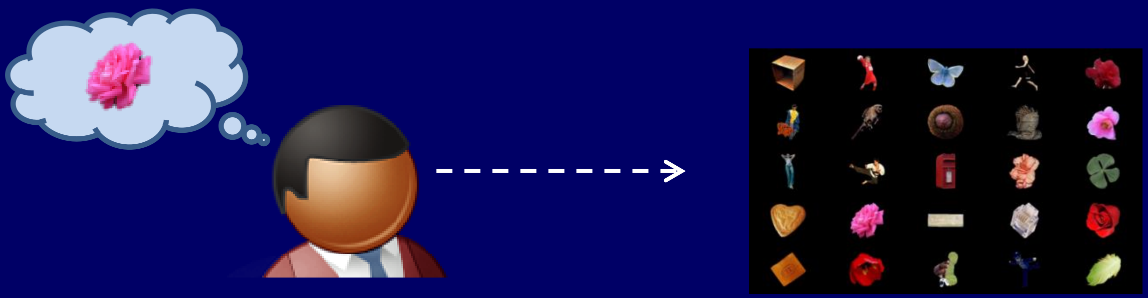
Human visual attention is important for designing gaze based systems.

Visual attention

- Bottom-up
  - When people view a scene with no intention

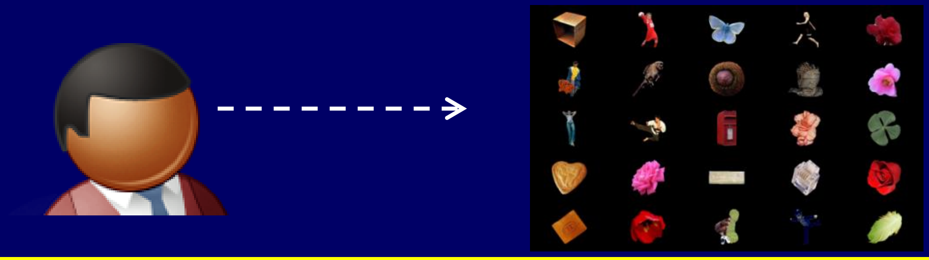


- Top-down
  - When people view a scene with intention

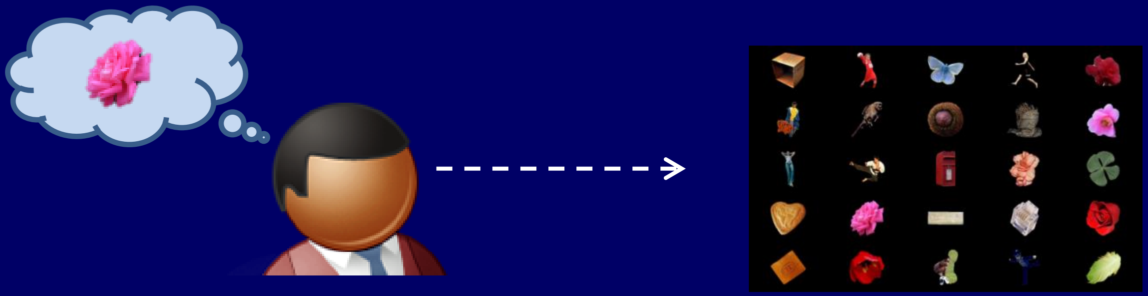


Visual attention

- Bottom-up
  – When people view a scene with no intention



- Top-down
  – When people view a scene with intention



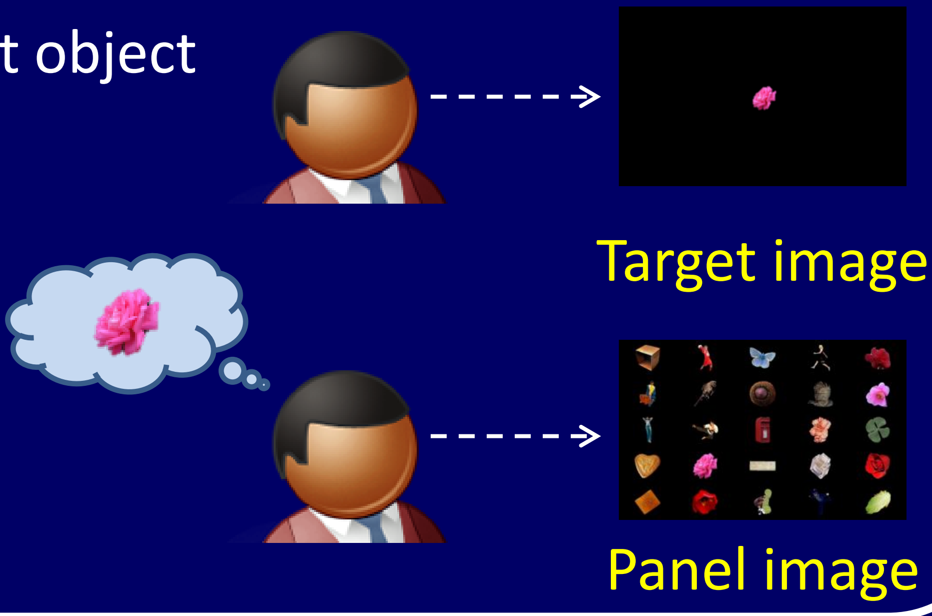
Purpose

- We estimate target-specific visual attention during the visual search task.

Visual search task

1.Memorize a target object

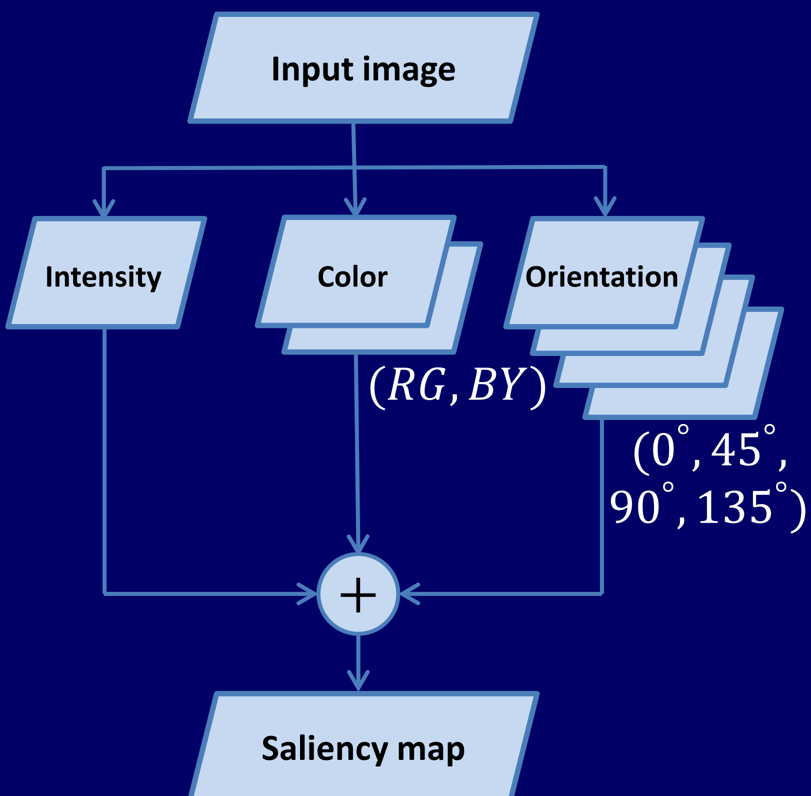
2.Search the target

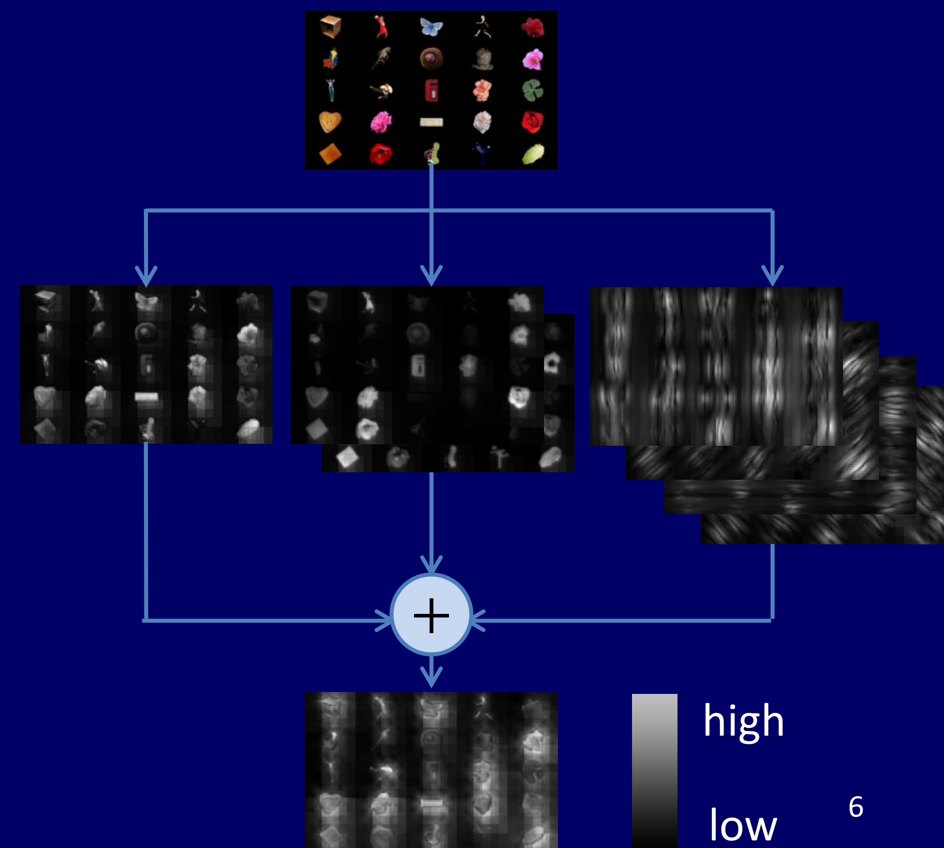


*We employed object images of MSRA Salient Object Database.

Related works

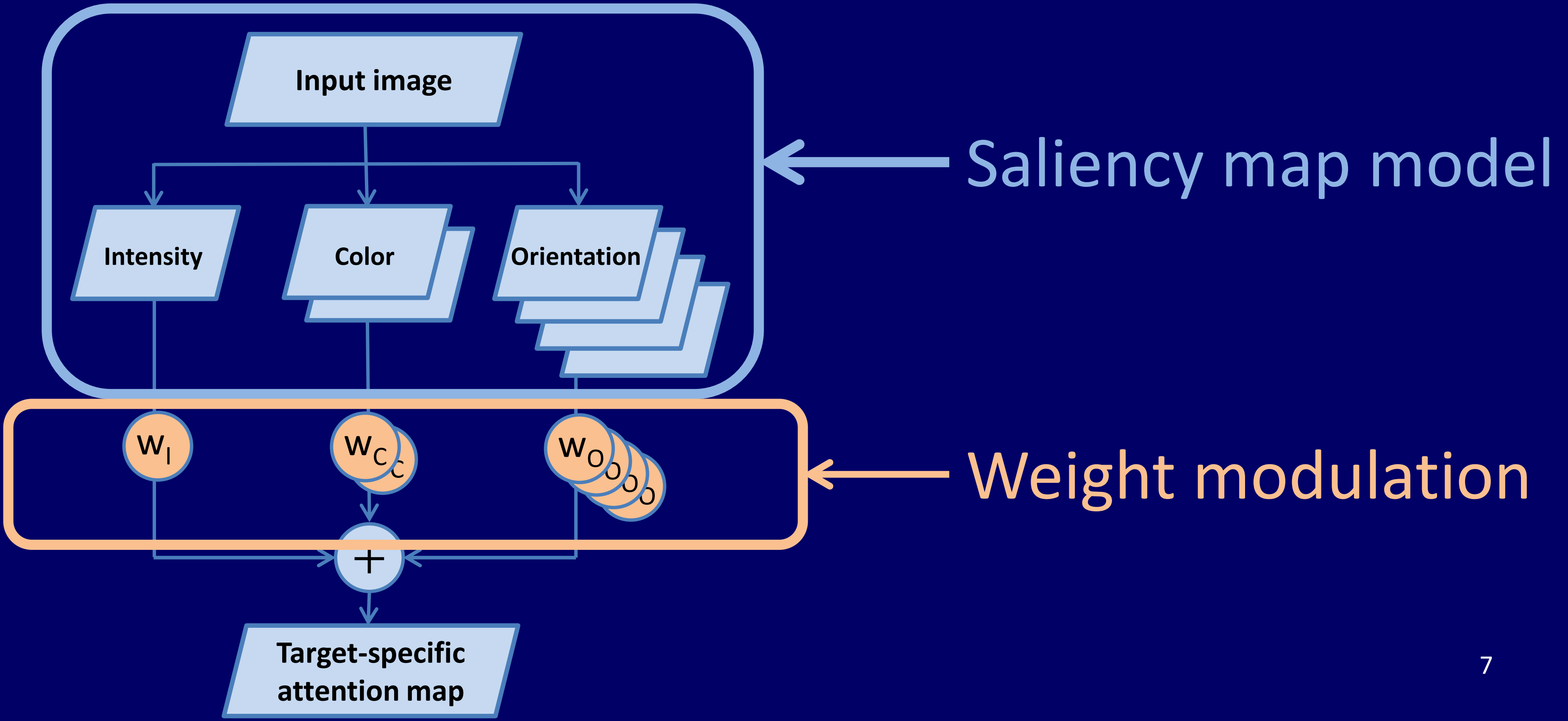
- Bottom-up visual attention estimation
  - Itti's Saliency map model
  - Only use input image





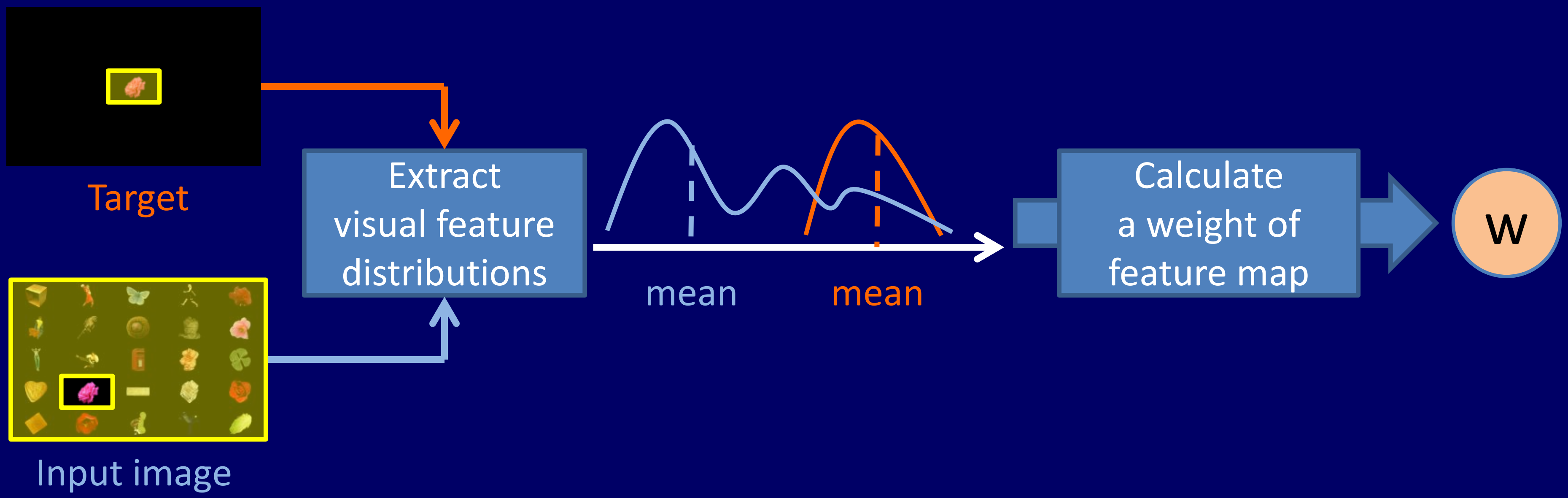
Related works

- Top-down visual attention estimation
  - Derive from Itti's Saliency map model
  - Use knowledge of target object

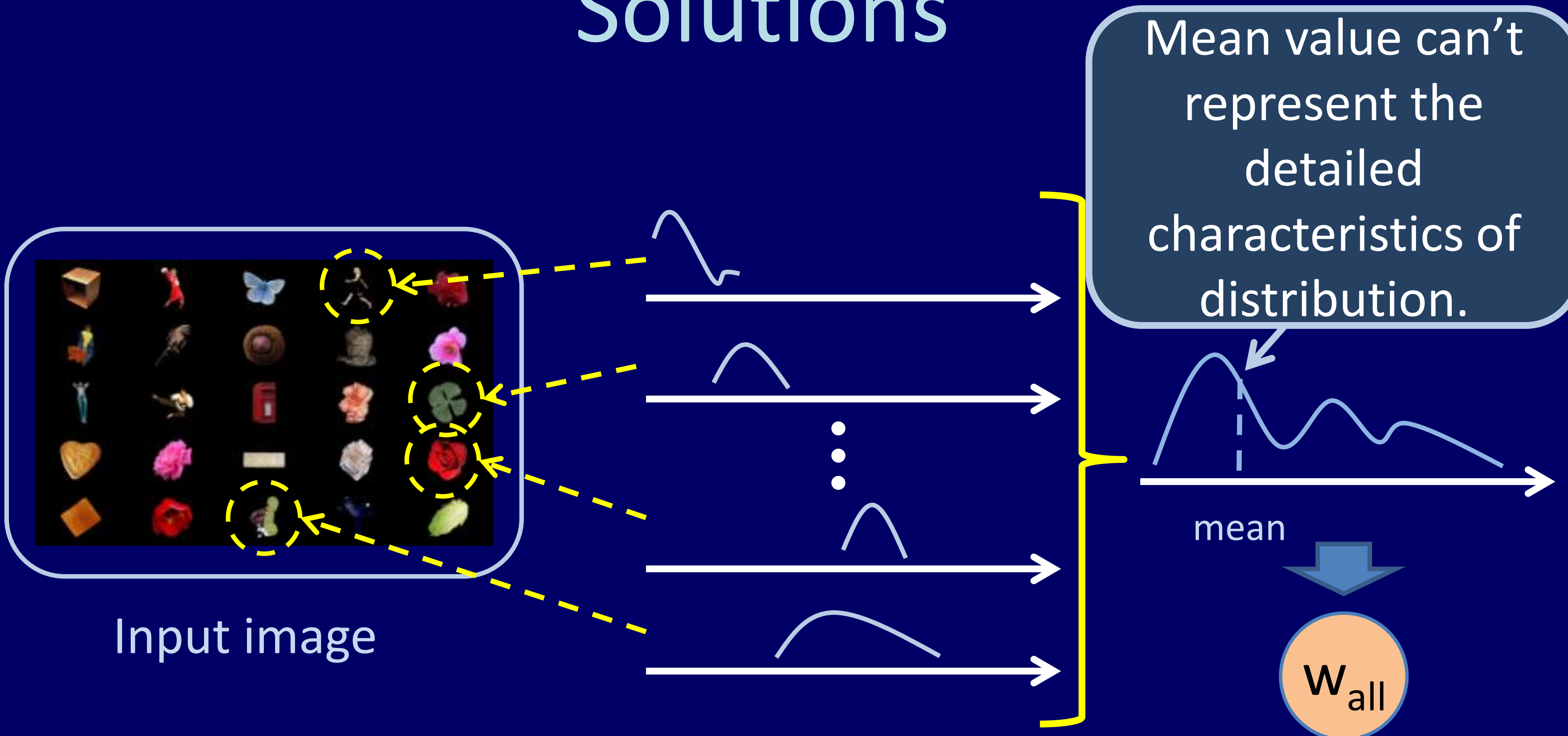


Related works

- Frintrop [2005], Navalpakkam [2006]
  - Consider a relationship between target and distractors



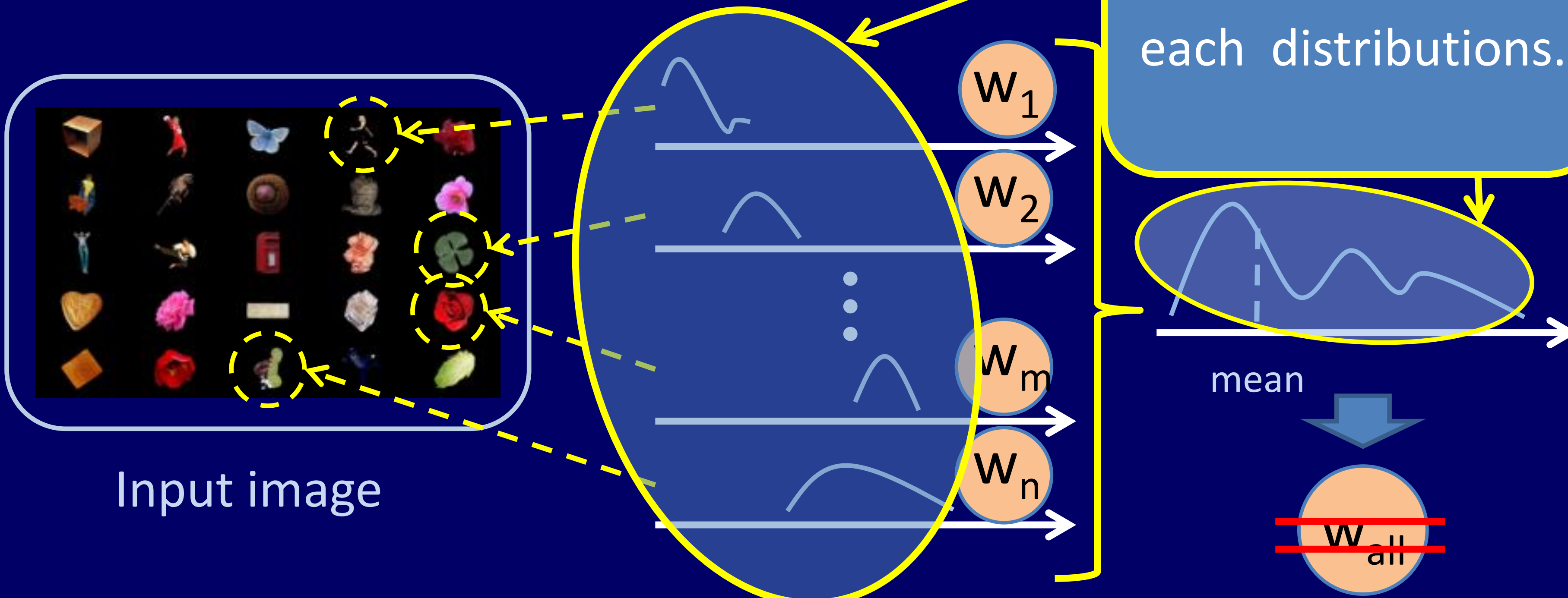
Solutions



It is difficult to estimate if input image contains complicated visual feature.

Approach

2. We consider each distributions.



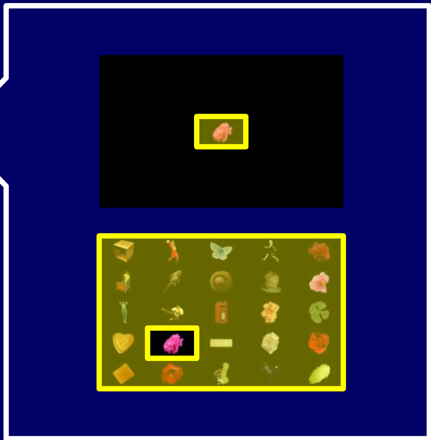
1. We focus on each object in panel image.
→Calculate spatially localized weights

Solutions

1. Calculate spatially localized weights

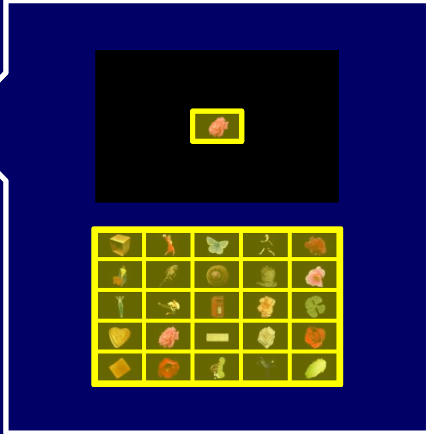
Related works

Relationship between target and all objects



Proposed

Relationship between target and each object

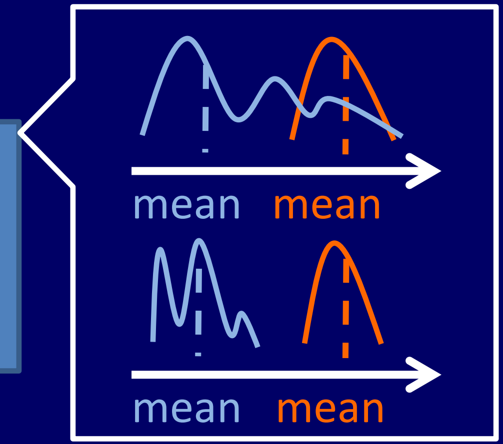


Solutions

2. Calculate the weights based on similarities between feature distributions

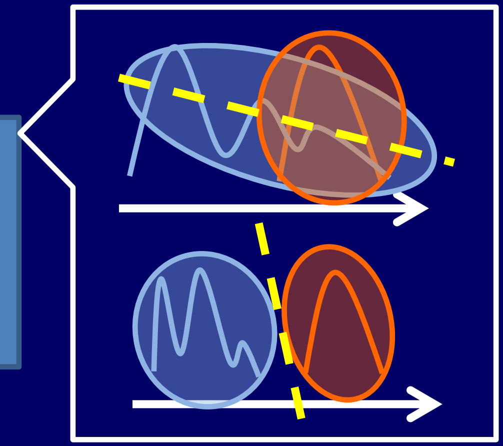
Related works

Relationship between mean value of visual features



Proposed

Linear separability of visual feature distributions



Linear separability of visual feature distributions

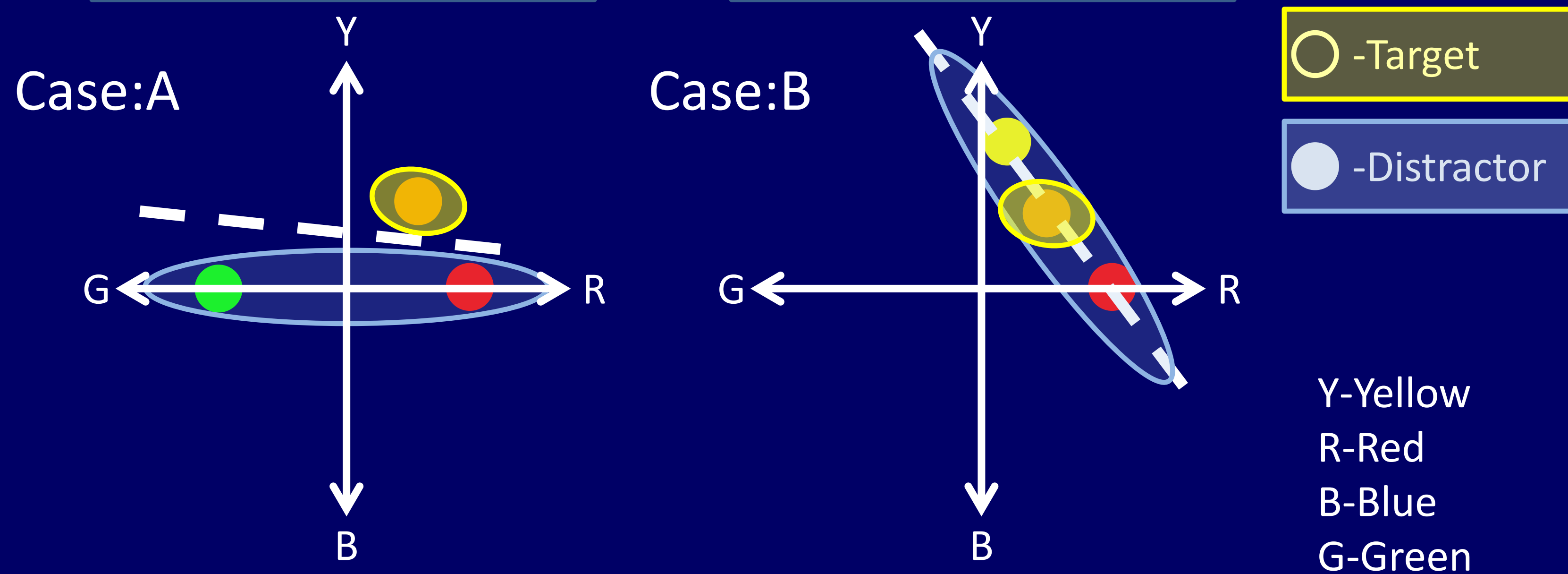
<Psychophysical findings>

Linear separability of visual feature distributions affects the performance of visual search.

Linear separability

Ex: Color feature[1]

Linearly separable	Nonlinearly separable



Variance ratio : High — easy to search

Variance ratio : Low — difficult to search

[1] John Hodsoll and Glyn W Humphereys, "Driving attention with top down: The relative contribution of target templates to the linear sepaprability effect in the size dimension", Perception and psychophysics, 63(5) , pp.918-926, 2001.

Linear separability

Ex: Color feature[1]

Linearly separable

Nonlinearly separable

Weight modulation of visual feature based on the inverse of variance ratio between the visual feature distribution of target and each object

Variance ratio : High

easy to search

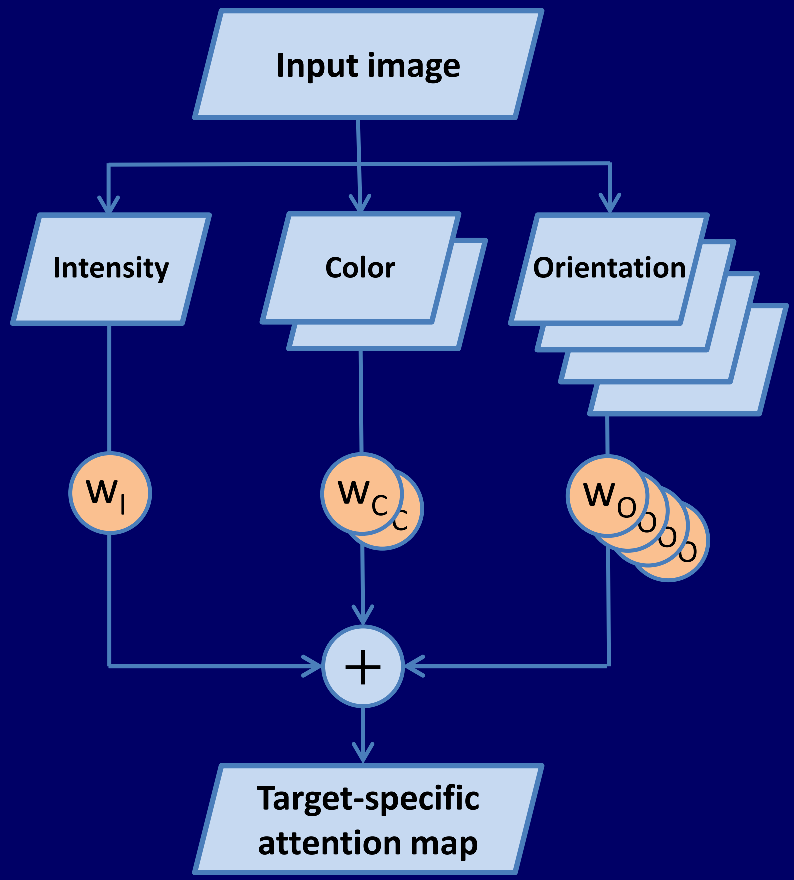
Variance ratio : Low

difficult to search

[1] John Hodsoll and Glyn W Humphereys, "Driving attention with top down: The relative contribution of target templates to the linear sepaprability effect in the size dimension", Perception and psychophysics, 63(5) , pp.918-926, 2001.

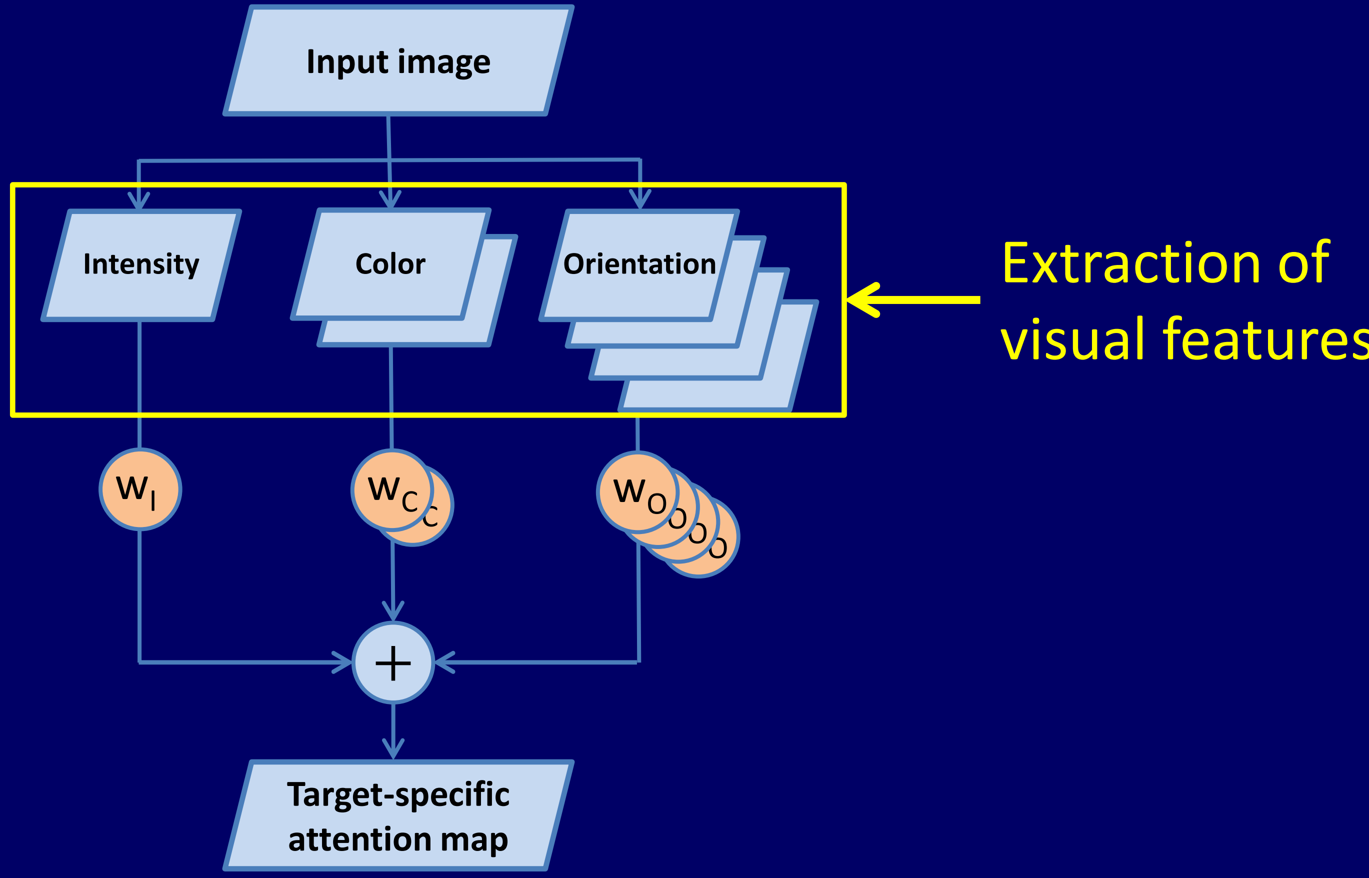
Proposed model

- Extention of Itti's bottom-up saliency map model



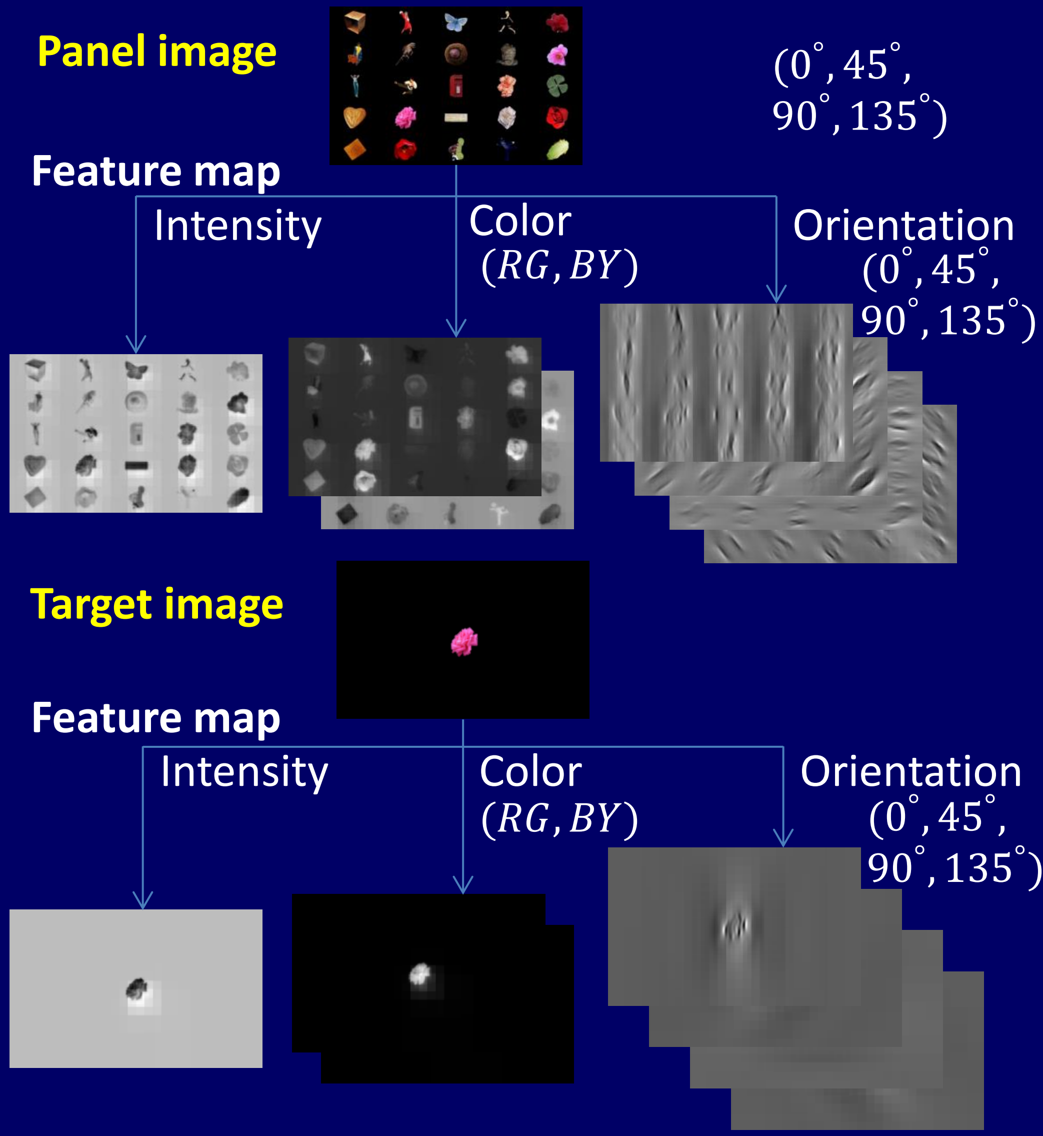
Proposed model(1/3)

- Extention of Itti's bottom-up saliency map model



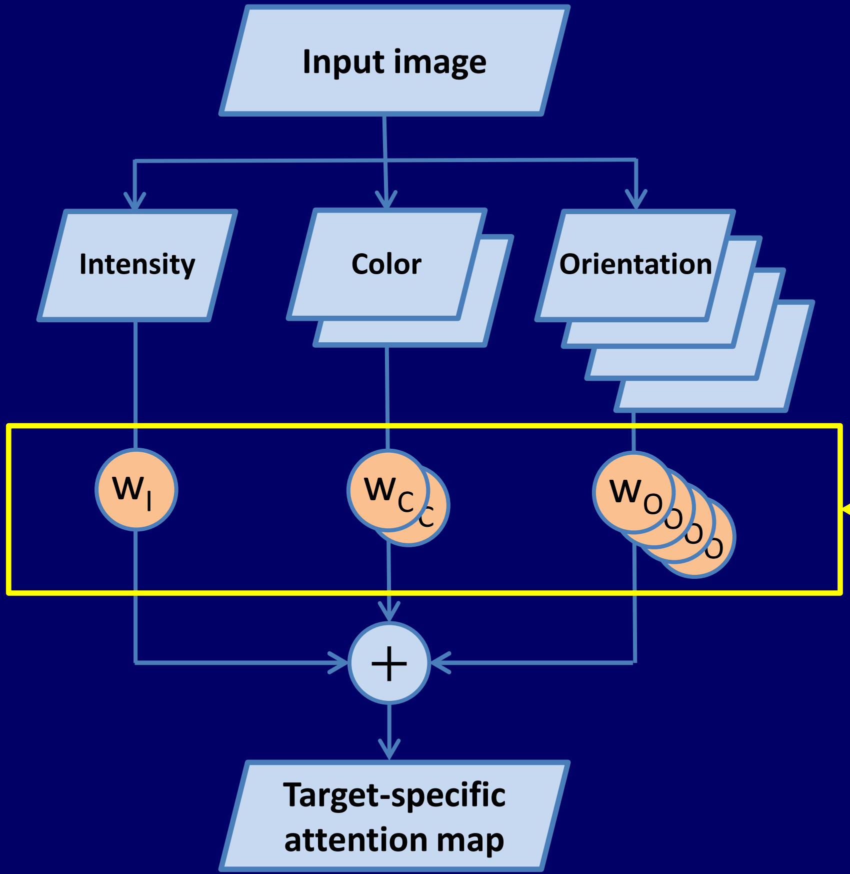
Extraction of visual features

- Extract visual features to create feature maps as with Itti's model



Proposed model(2/3)

- Extention of Itti's bottom-up saliency map model



Weight modulation

Region segmentation

- Weight modulation based on the relationships between target and each object

Segment each feature map into equal sub regions

Ex. Color RG

Tile region

Weight modulation

- Modulate the weight of feature maps based on the linear separability of visual search
  →Use the Fisher's variance ratio (J)

Low J→The object is similar to the target
→Give higher weight

High J→The object is not similar to the target
→Give lower weight

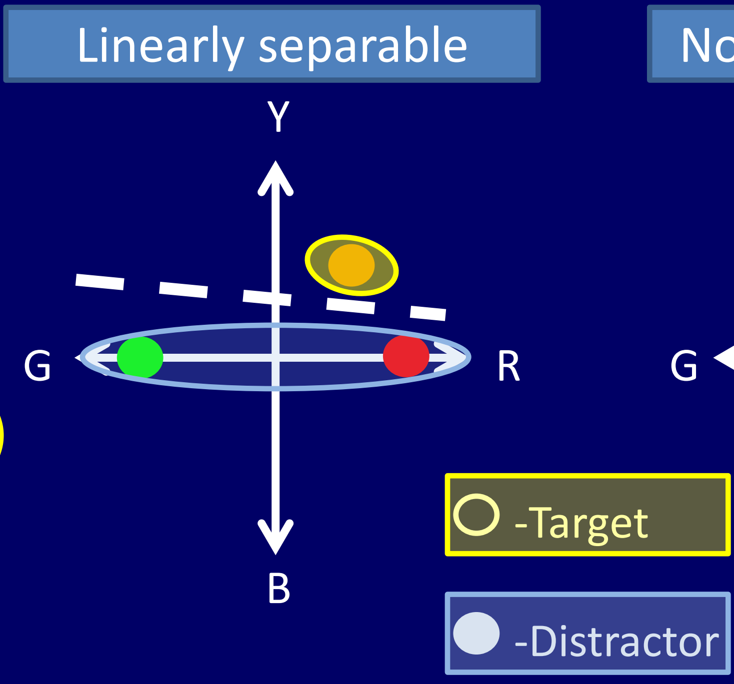
$$w = \frac{1}{J}$$

Weight modulation

- Modulate the weight of feature maps based on the linear separability of visual feature distributions

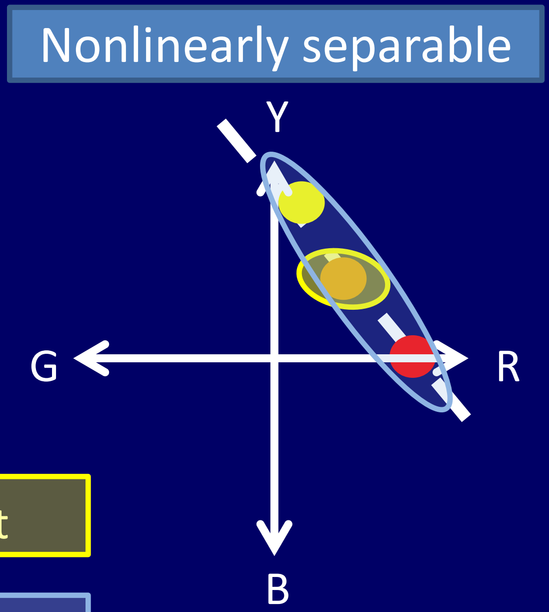
Fisher's variance ratio (J)

$$w = \frac{1}{J}$$



J : High

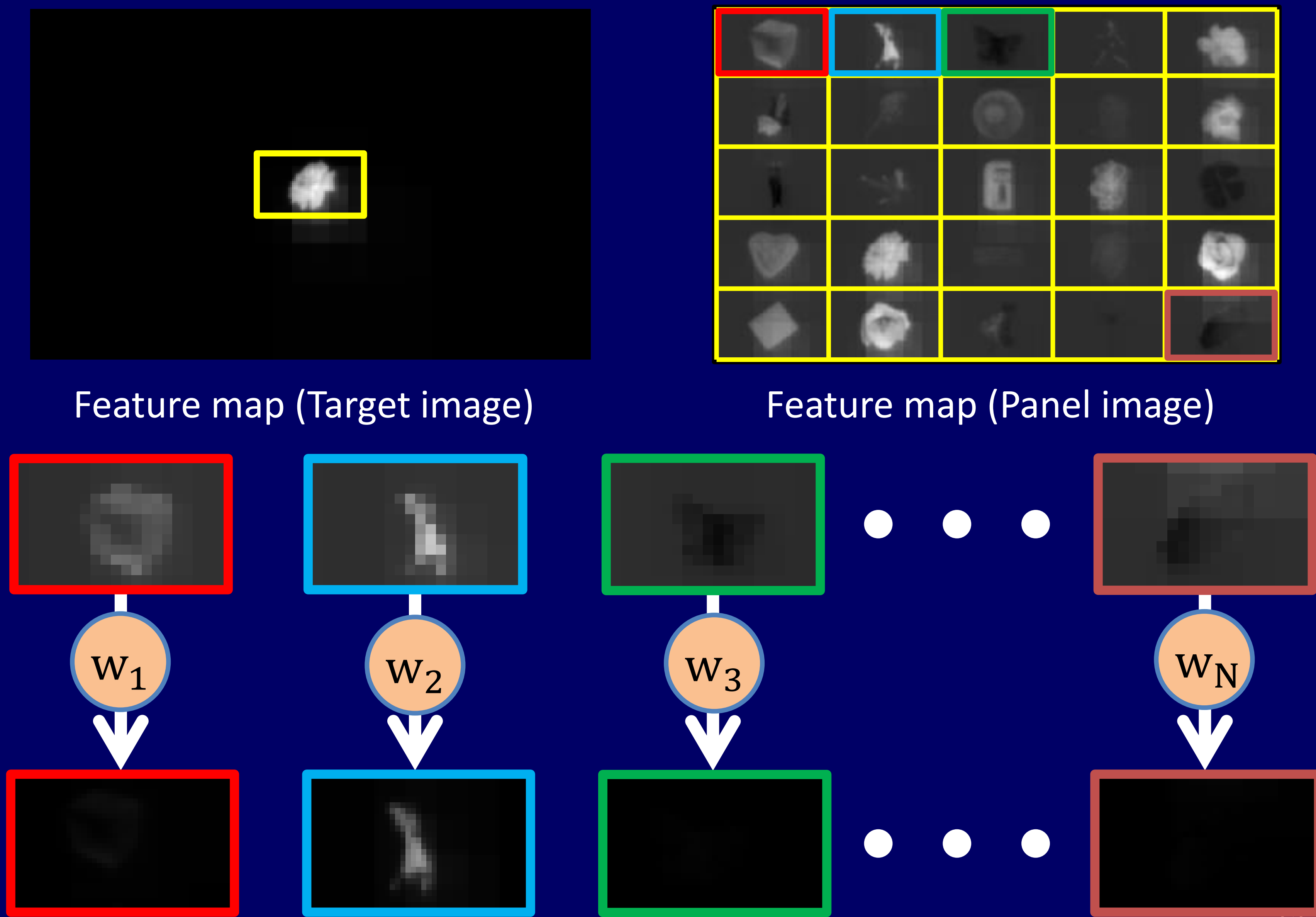
Not similar to the target
→Give lower weight



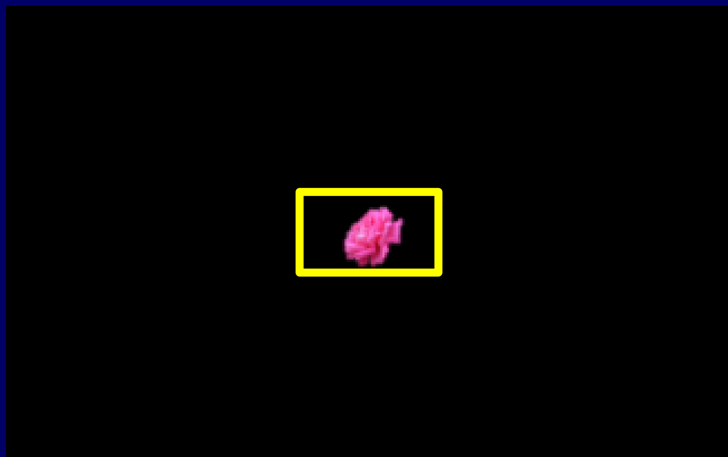
J : Low

Similar to the target
→Give higher weight

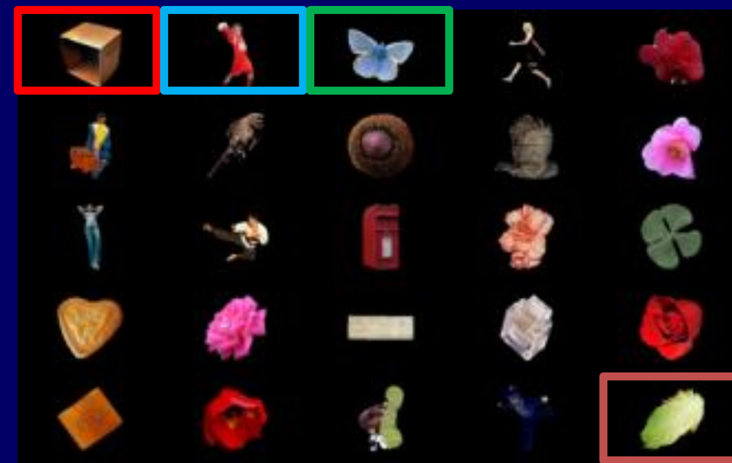
Calculation of spatially localized weights



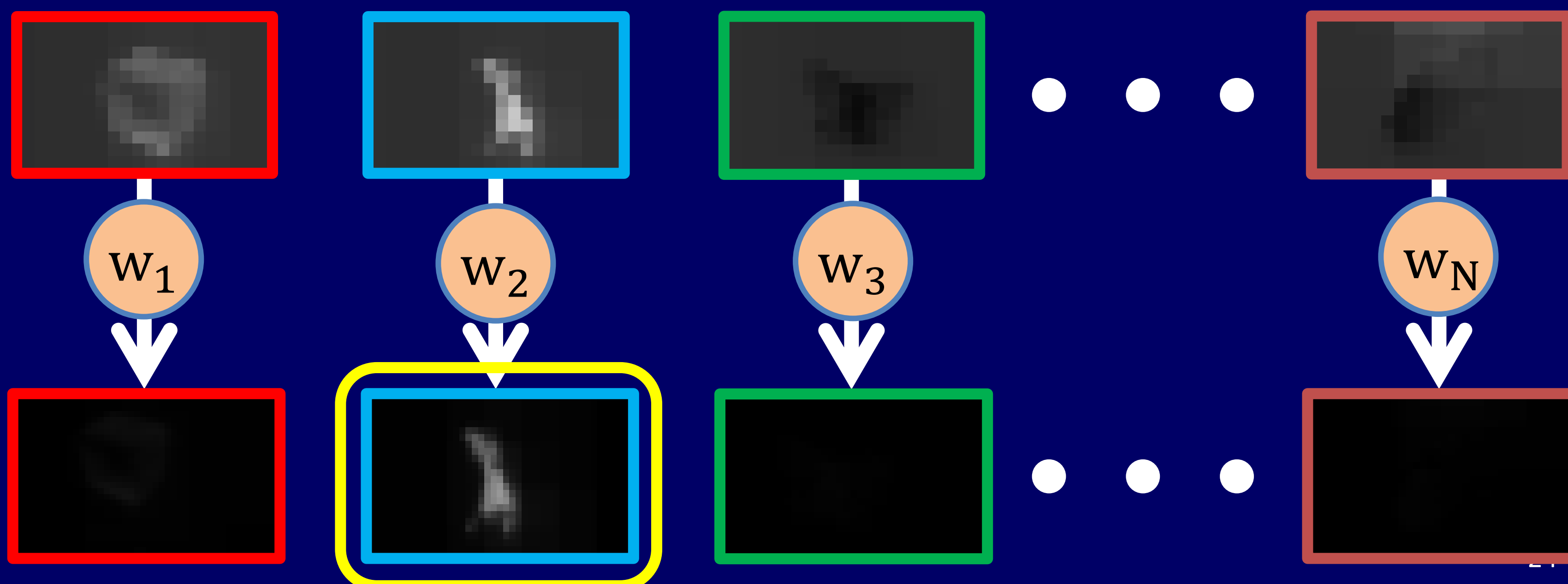
Calculation of spatially localized weights



Feature map (Target image)

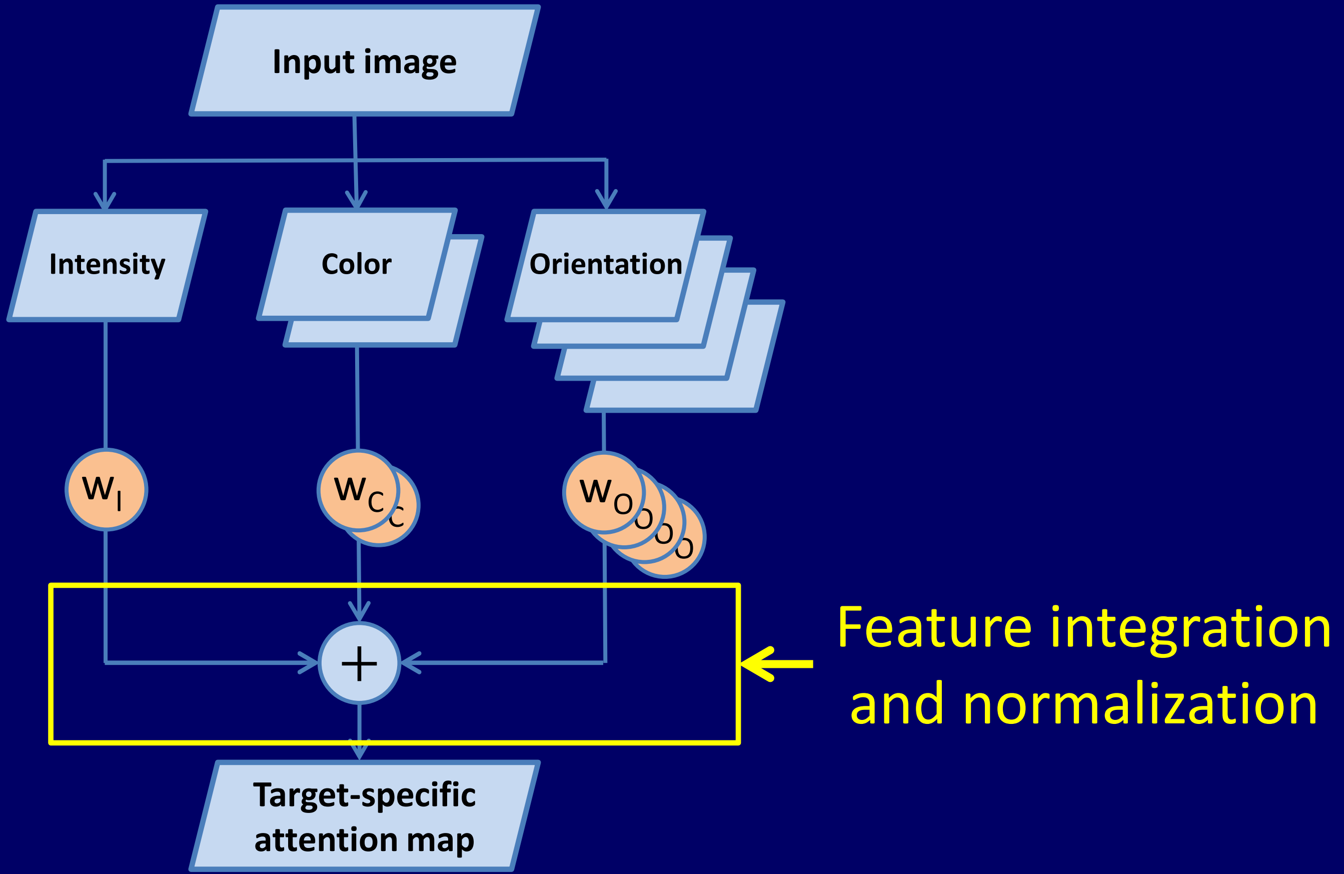


Feature map (Panel image)



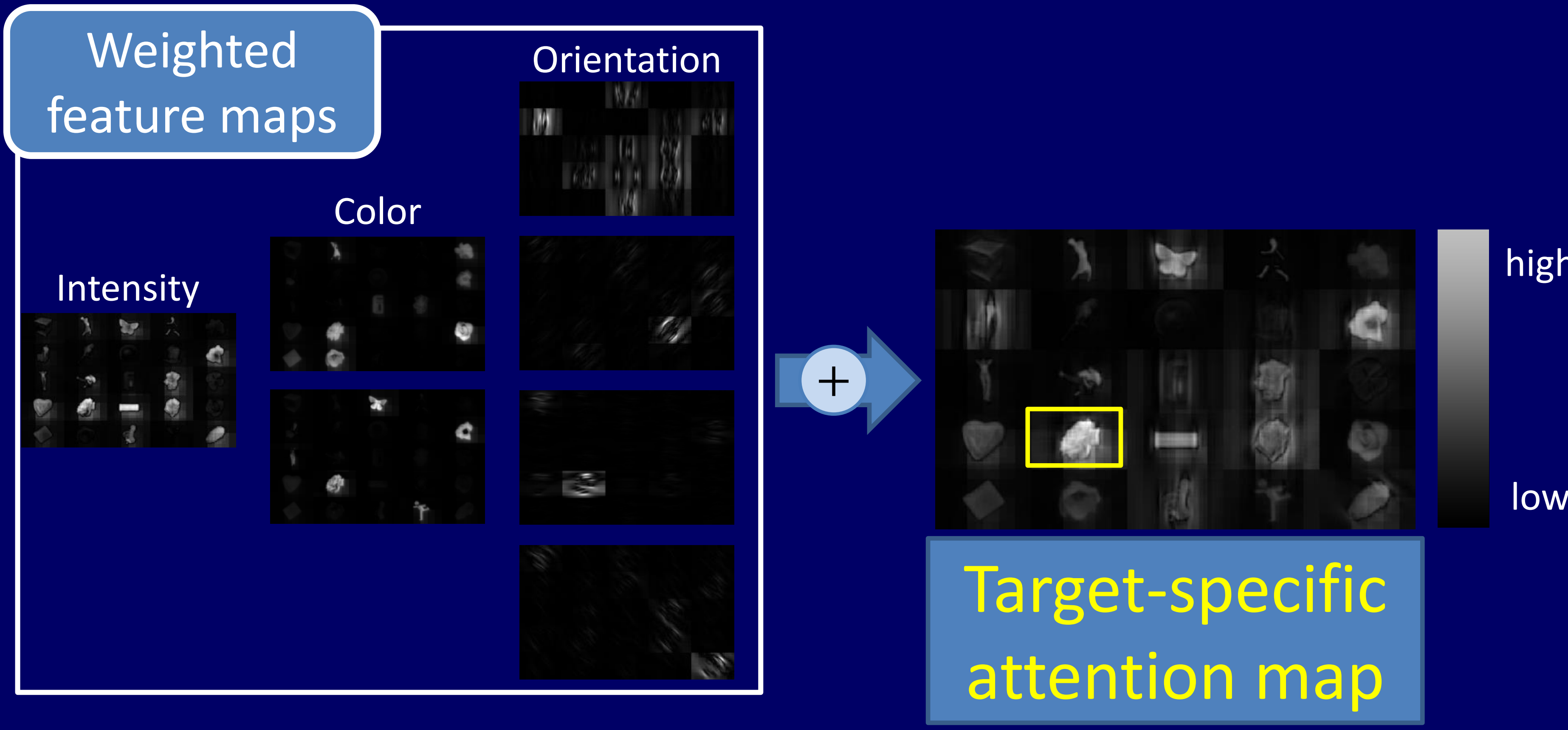
Proposed model(3/3)

- Extension of Itti's bottom-up saliency map model



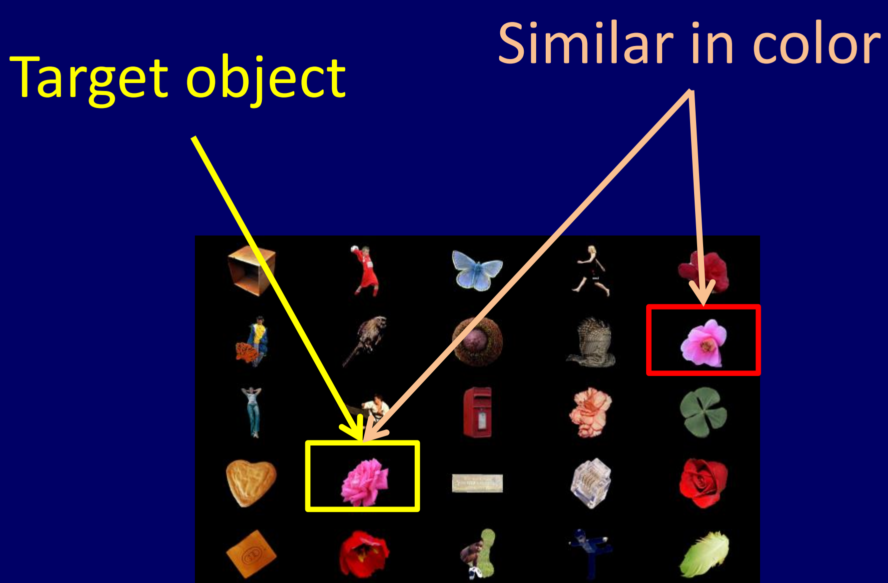
Feature integration and normalization

- Integrate the weighted seven feature maps into a target-specific attention map and normalize it



Feature integration and normalization

- Integrate the weighted seven feature maps into a target-specific attention map and normalize it

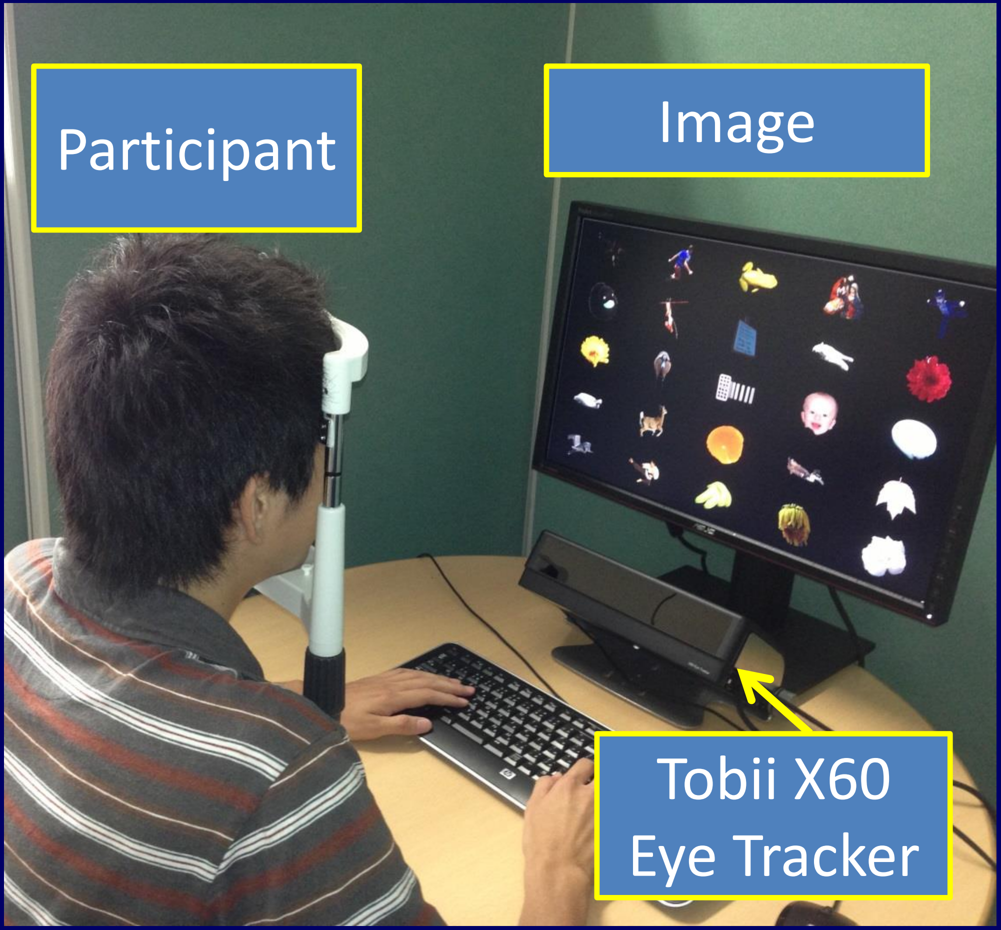


Input image

Target-specific attention map

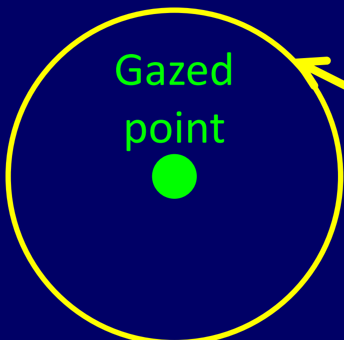
Experiment evaluation

- Measure gaze data during visual search
- Participant:10 people （Male：9, Female：1）
  - 100 trials × 10 people
  - 740 fixation sequences



Gaze data

- Calculate gazed areas from gaze points



Gazed area

Error range of eye tracker(visual angle:1.0°)

\+

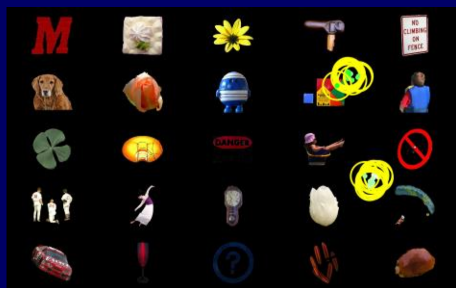
Visual range of central fovea(visual angle:2.0°)

Gazed points (fixation data)

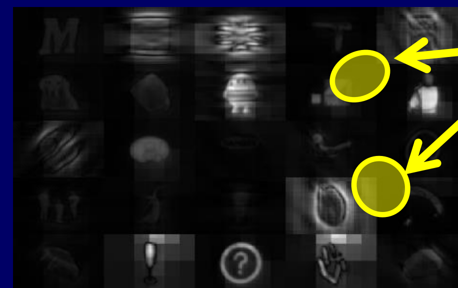
Gazed areas

Measures for approach

- Normalized Scanpath Saliency(NSS)
  - Response value at the gaze point on the attention map



Gazed areas



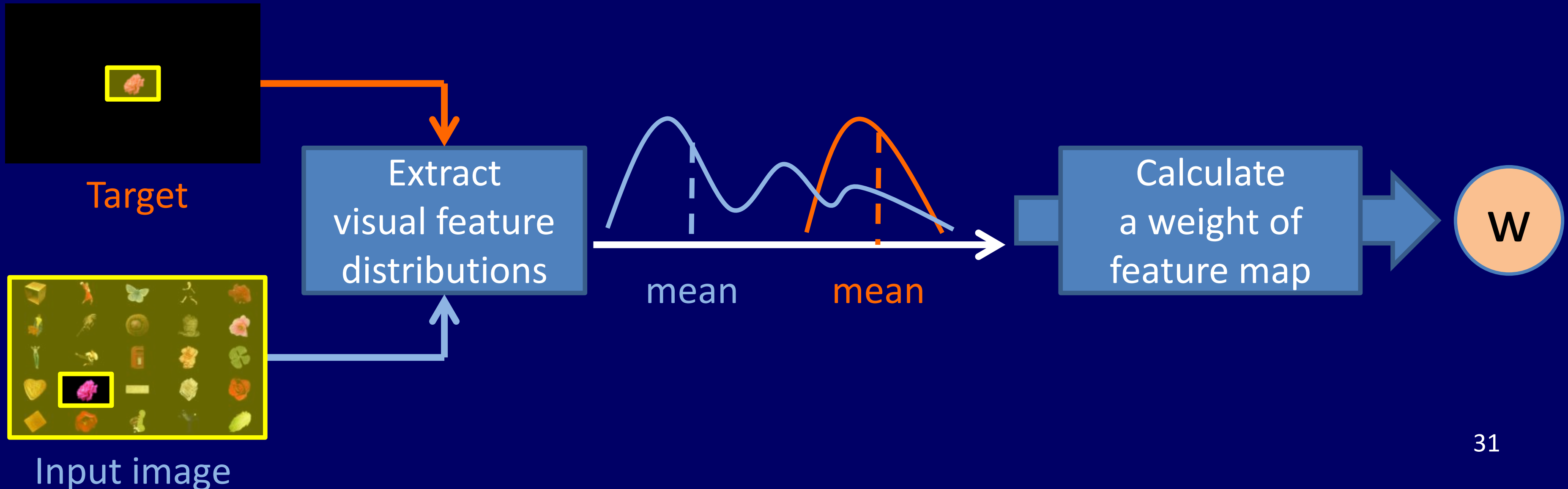
Attention map

NSS:
average value

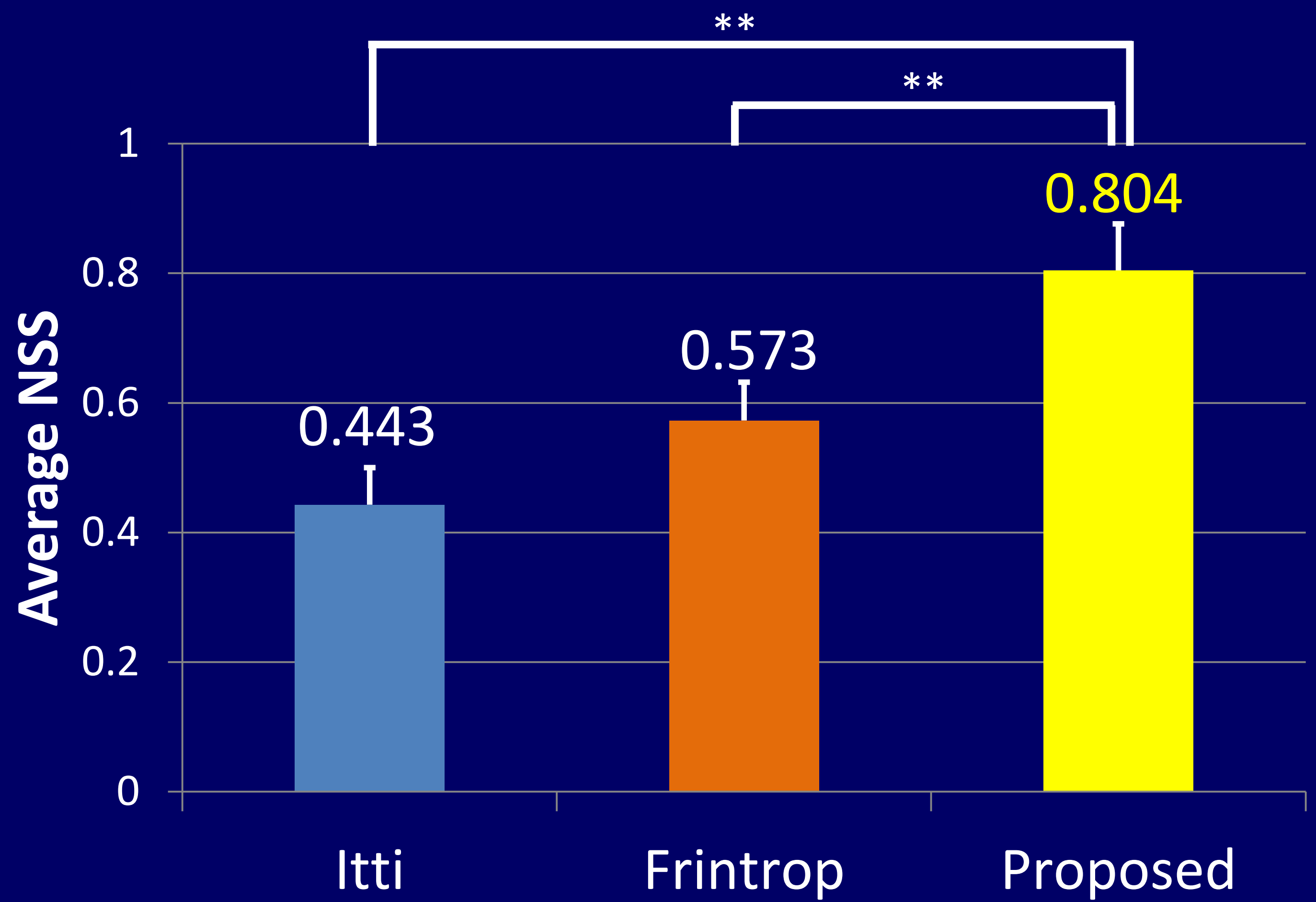
- Top measurements based on NSS
  - Ave. NSS : Average of NSS at gazed areas
  - Ave. NUSS : Average of the response values except at the gazed areas to evaluate false detection
- Comparative model
  - Itti’s model and Frintrop’s model

A comparative model

- Frintrop's top-down visual attention computational model
  - Modulation the weight based on relationship between target and all objects
  - Apply the weight to the overall feature map

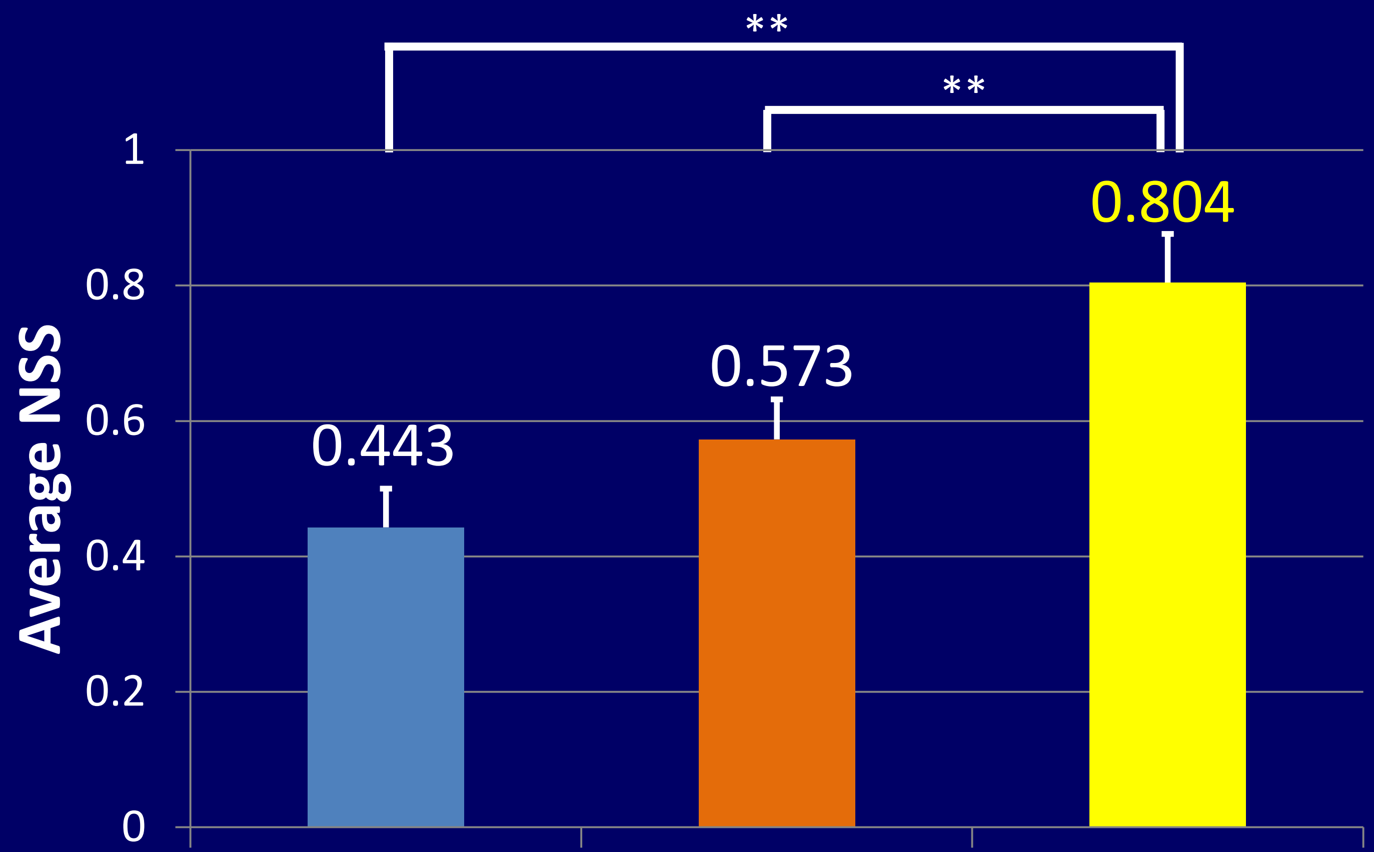


Result(Average NSS)



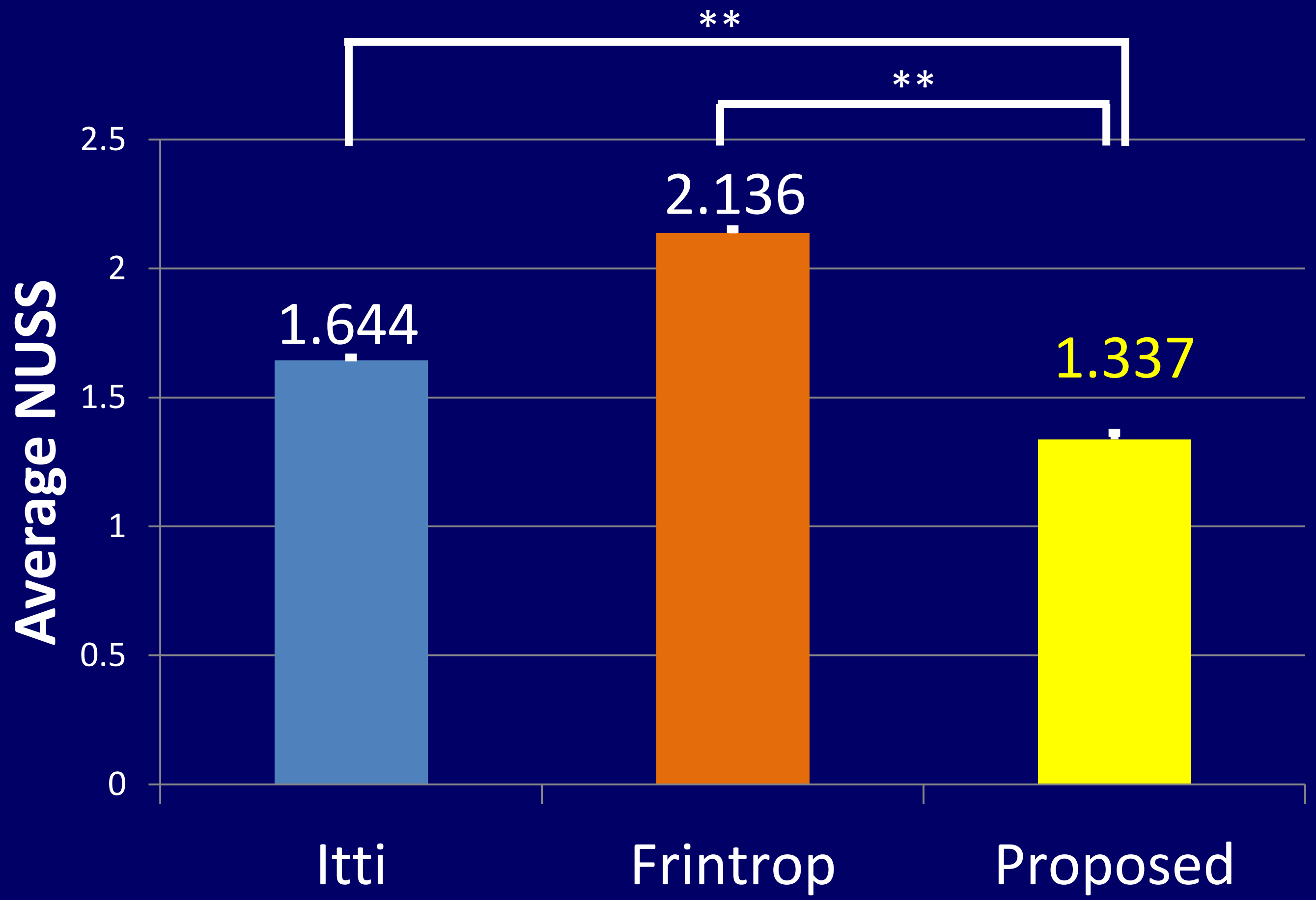
$** : p < 0.01$

Result(Average NSS)



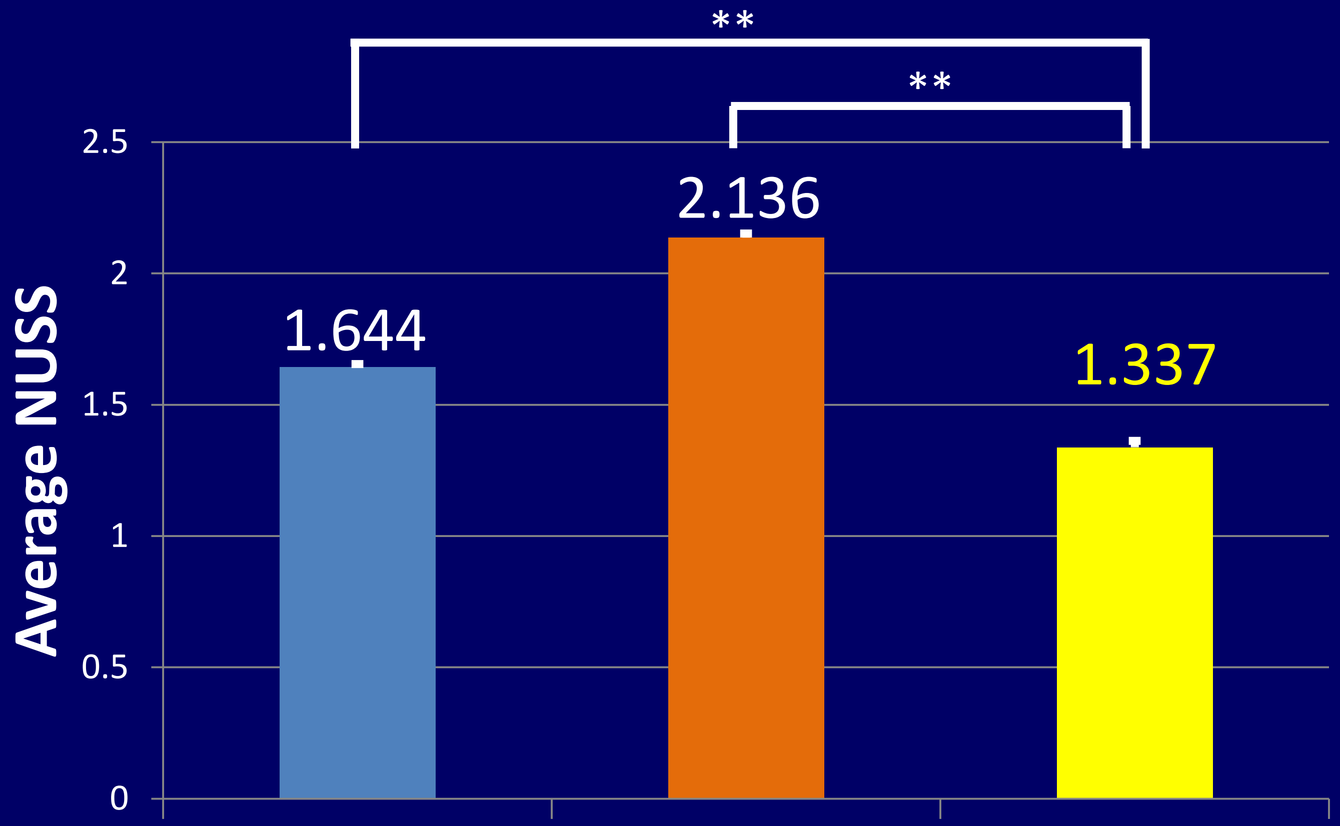
Higher average NSS
→Our model can estimate actual focused areas

Result(Average NUSS)



$** : p < 0.01$

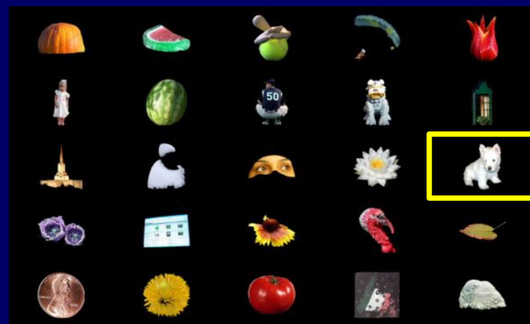
Result(Average NUSS)



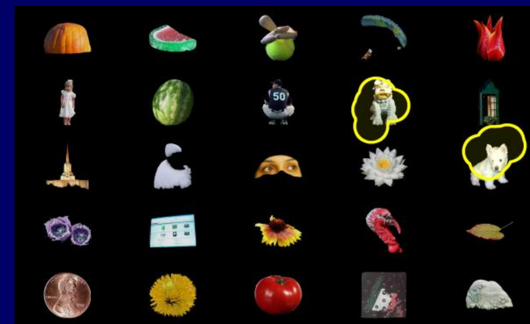
Lower average NUSS

→Our model can suppress false detection

Higher NSS in our model



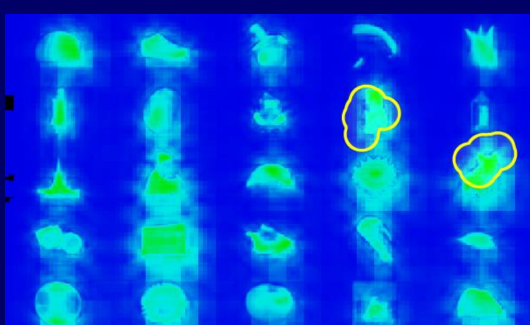
Panel image



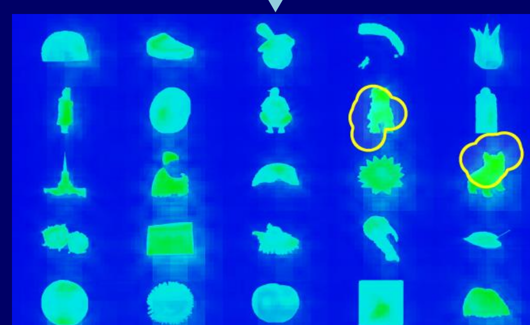
Gazed areas

High response except at gazed areas
(High false detection rate)

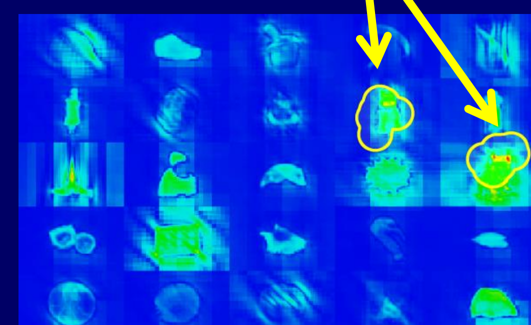
High response at gazed areas
(High detection rate)



Itti

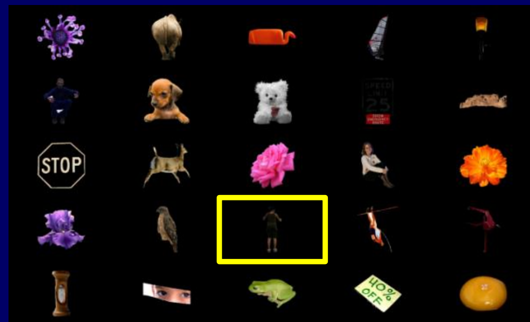


Frintrop



Proposed

Lower NSS in our model

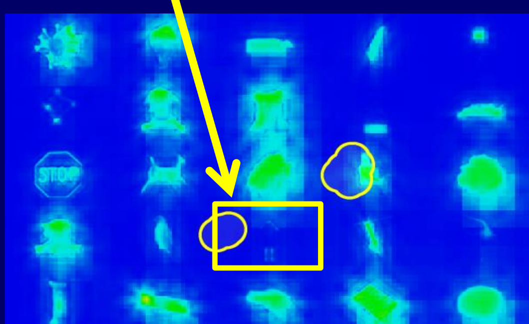


Panel image

Gazed areas

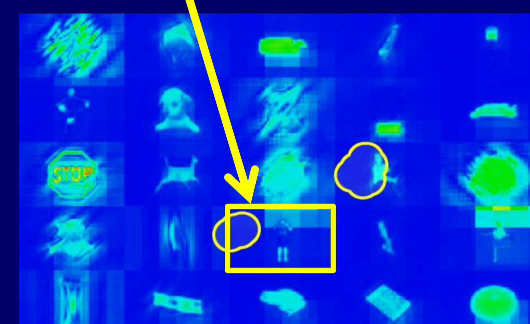
Low bottom up saliency
at target area

Low response at the target area
(Low detection rate)



Itti

Frintrop



Proposed

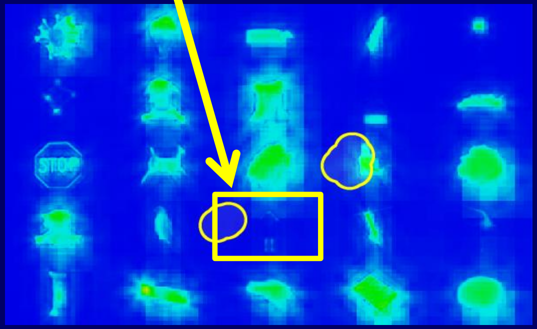
Lower NSS in our model

Panel image

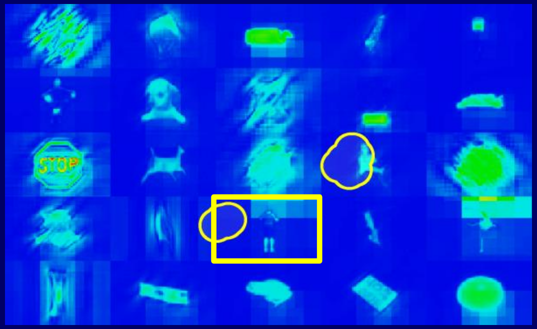
Gazed areas

Low bottom up
saliency at the target area

High response
at the target area


Itti


Frintrop


Proposed

Conclusions

- Target-specific visual attention computational model.
  - Extension of Itti's bottom-up saliency map model
  - Application of psychophysical findings on visual search to weight modulation of visual feature map
- High estimation accuracy of visual attention
  - High normalized scanpath saliency (NSS)
  - Less false detection

Future works

- Evaluate the sequence of scan path
- Design a generalized model without region segmentation
- Verify our model using natural images