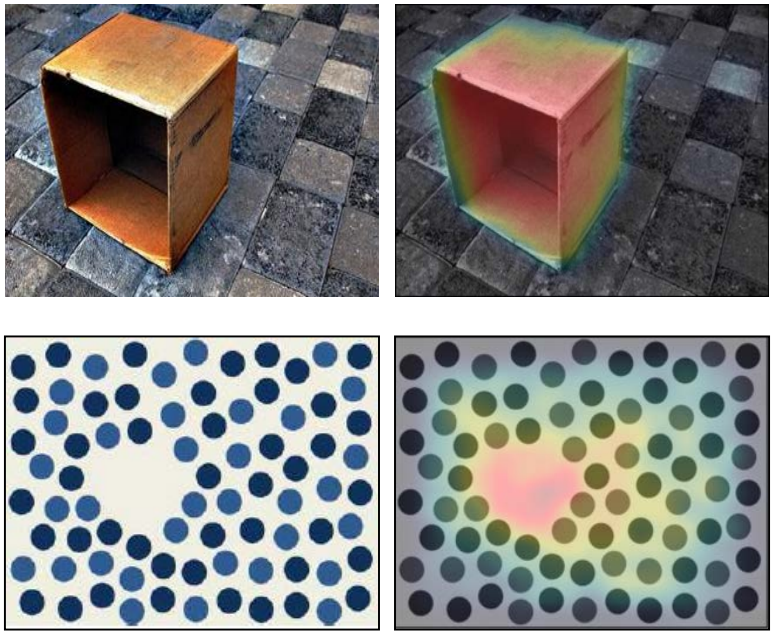
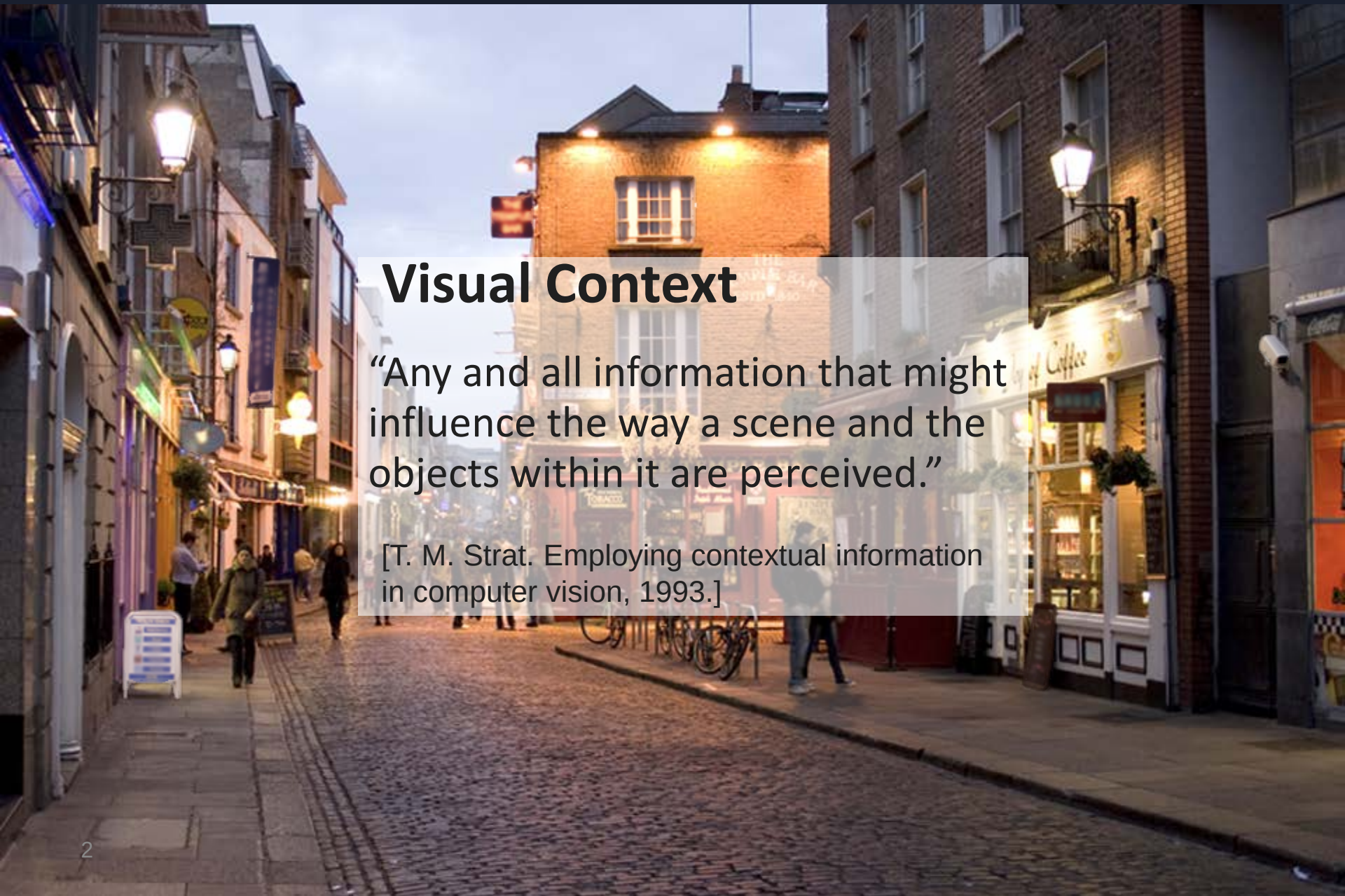
A Closer Look at Context: From Coxels to the Contextual Emergence of Object Saliency



Rotem Mairon and Ohad Ben-Shahar
Computer Science Department, Ben-Gurion University of the Negev.

Interpretations of Visual Context



Visual Context

"Any and all information that might influence the way a scene and the objects within it are perceived."

[T. M. Strat. Employing contextual information in computer vision, 1993.]

Visual Context in Salient Objects Detection

Saliency:

The degree to which a constituent in a visual scene demands our attention.

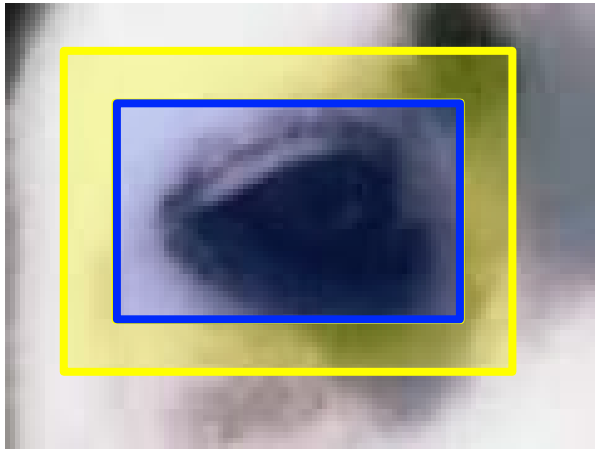
Visual Context in Salient Objects Detection

Visual context of a constituent:

The set of visual units that are used to measure its saliency.

Context in previous methods:

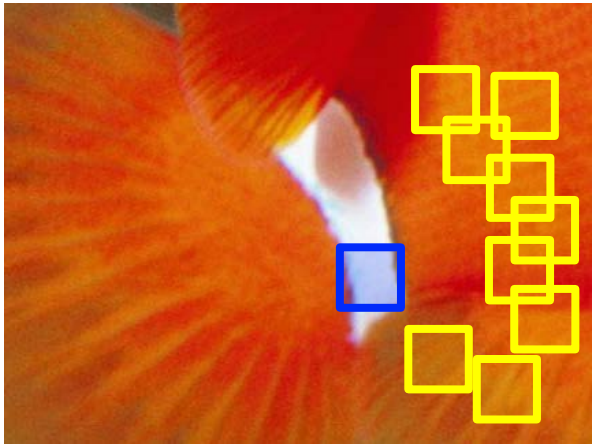
Contrast



Context = the surround in some center-surround (C-S) structure.

Wang et al. 2008, Liu et al. 2011, …

Rarity



Context = a set of patches derived from global image statistics.

Achanta et al. 2009, Margolin et al. 2013, …

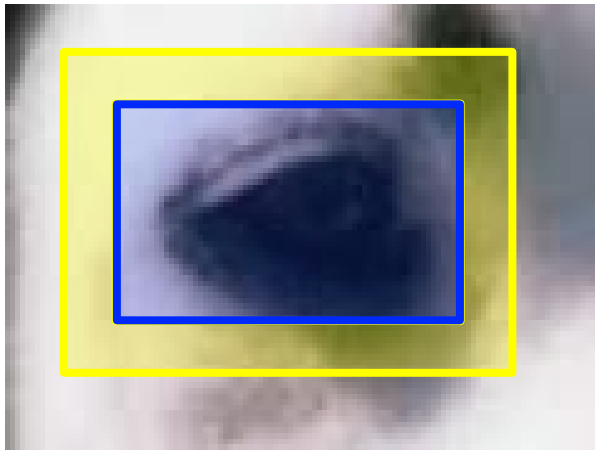
Visual Context in Salient Objects Detection

Visual context of a constituent:

The set of visual units that are used to measure its saliency.

Context in previous methods:

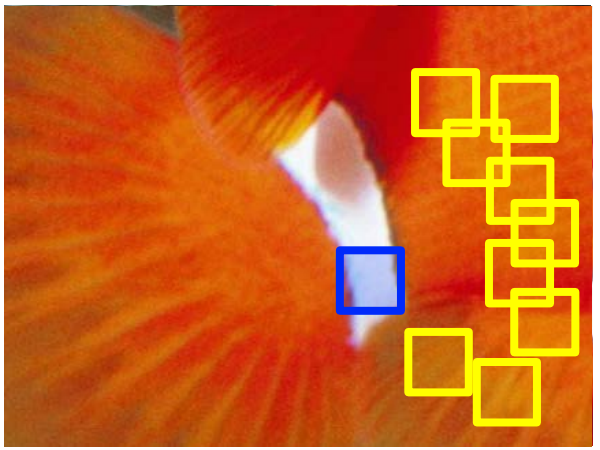
Contrast



Context = the surround in some center-surround (C-S) structure.

Wang et al. 2008, Liu et al. 2011, …

Rarity

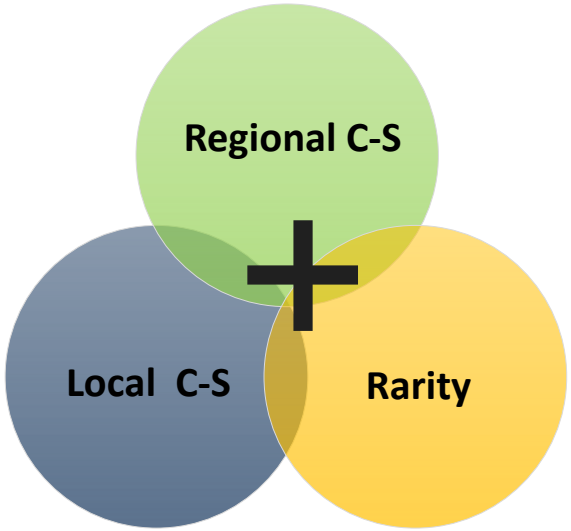


Context = a set of patches derived from global image statistics.

Achanta et al. 2009, Margolin et al. 2013, …

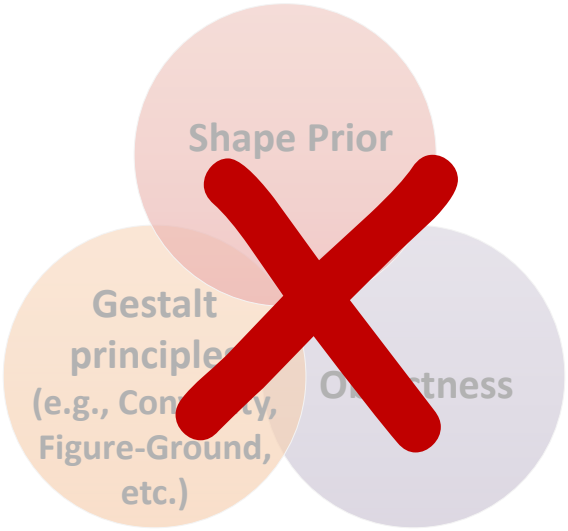
Visual Context in Salient Objects Detection

Context in previous methods



Not fully effective in measuring the saliency of whole visual objects.

Additional information

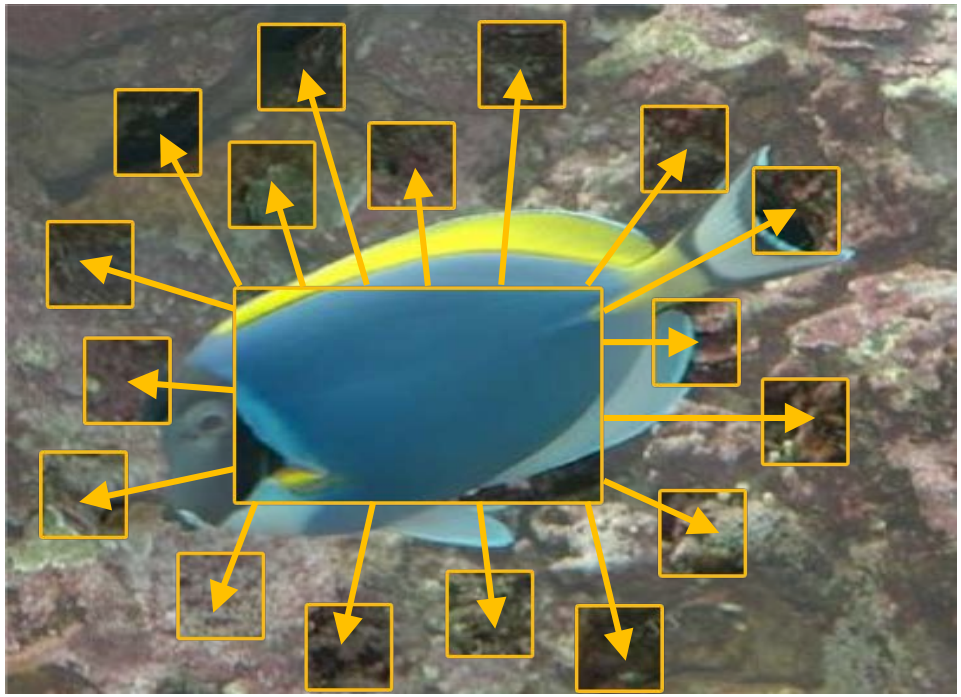


Hu et al. 2004, Chang et al. 2011, Jiang et al. 2011, Wang et al. 2011, …

Our approach: Improve performance by better modelling context instead of objects.

Visual Constituent VS Visual Context

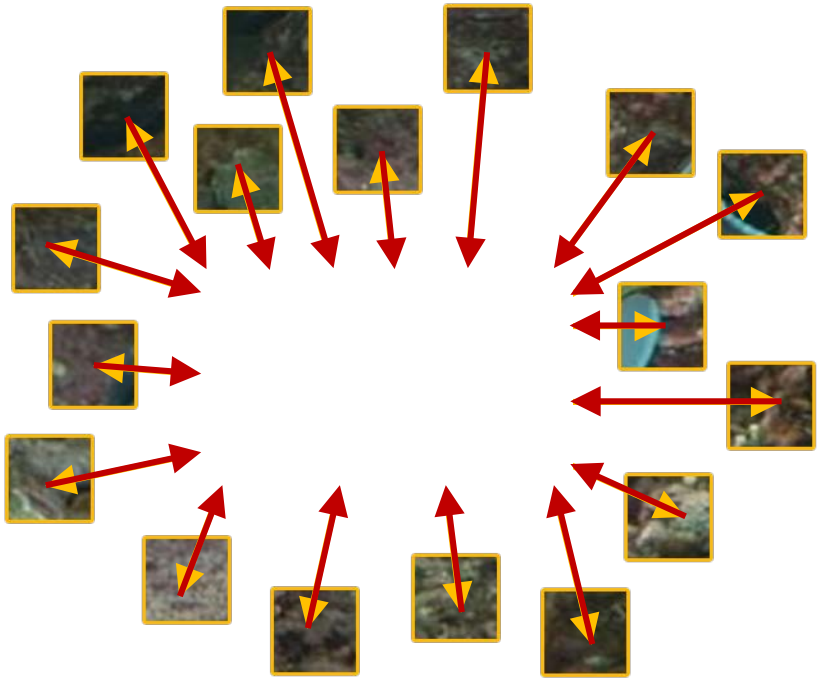
"To what extent does a visual **constituent** **stand out** from its **context**?"



- Implies a certain constituent is at hand when its saliency is measured.
- When the **constituent** is intended to be an **object**, performance is limited.

Visual Constituent VS Visual Context

"What are the properties of **visual context** that render the **visual information it embeds** as salient?"



- Modelling visual context instead of the object
- No need to consider object shape/size/complexity
- "Non-contextual" information is indicated as salient.

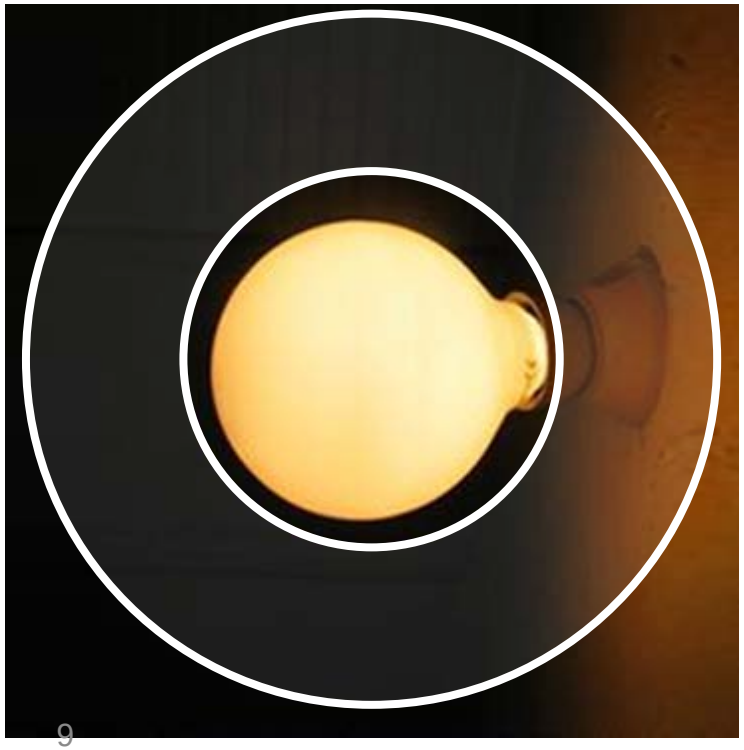
A Novel Model of Visual Context

Context Element (coxel): A region in the image with the following properties:

Smoothness: Proximate parts composing the coxel are expected to be similar.

Apathy to contiguity:

Enclosure:



A Novel Model of Visual Context

Context Element (coxel): A region in the image with the following properties:

Smoothness

Apathy to contiguity: A coxel may constitute several distinct connected components.

Enclosure:

A Novel Model of Visual Context

Context Element (coxel): A region in the image with the following properties:

Smoothness

Apathy to contiguity

Enclosure: The spatial layout of a coxel should “enclose” some visual information.

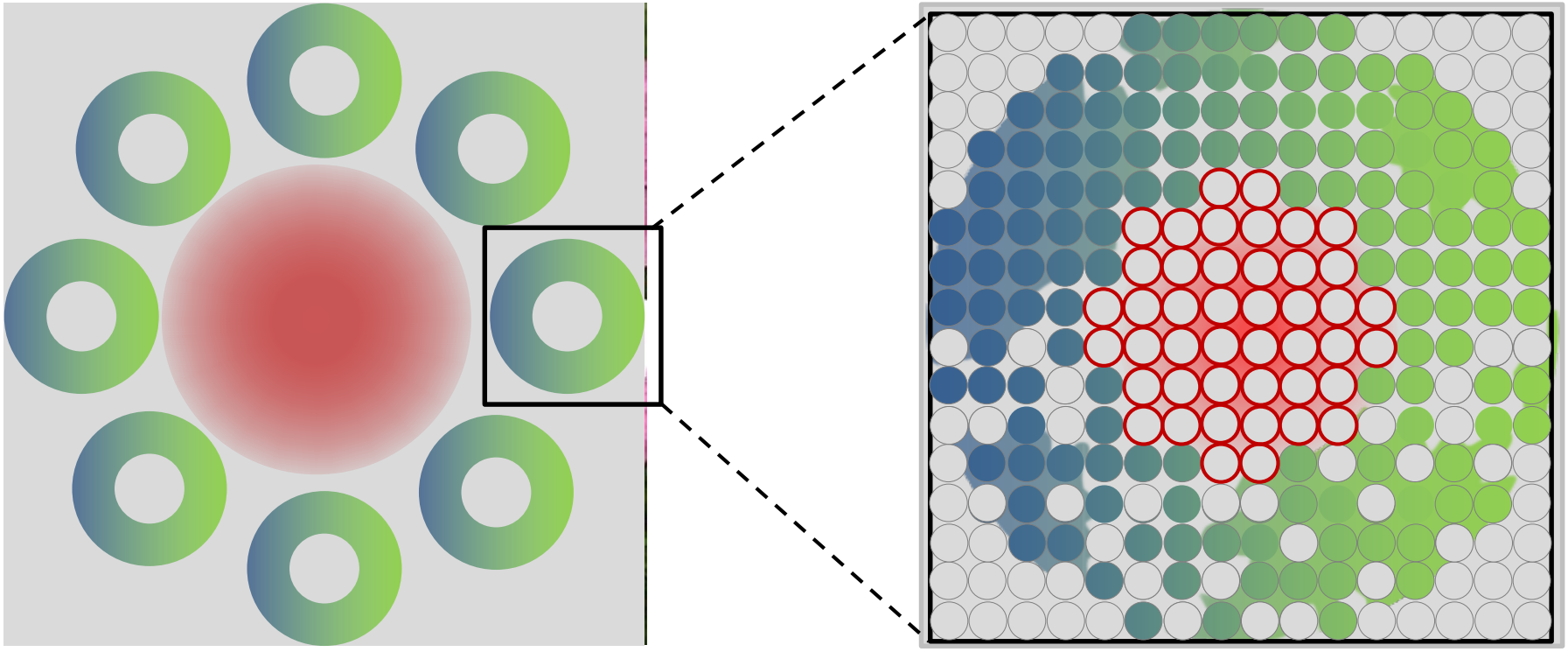
A Novel Model of Visual Context

Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

Smoothness
- $V=\{v_1 \ldots v_n\}$: the set of all patches in the input image.

Apathy to contiguity
- $G=(V,E)$: a *weighted complete* graph on V.

Enclosure:

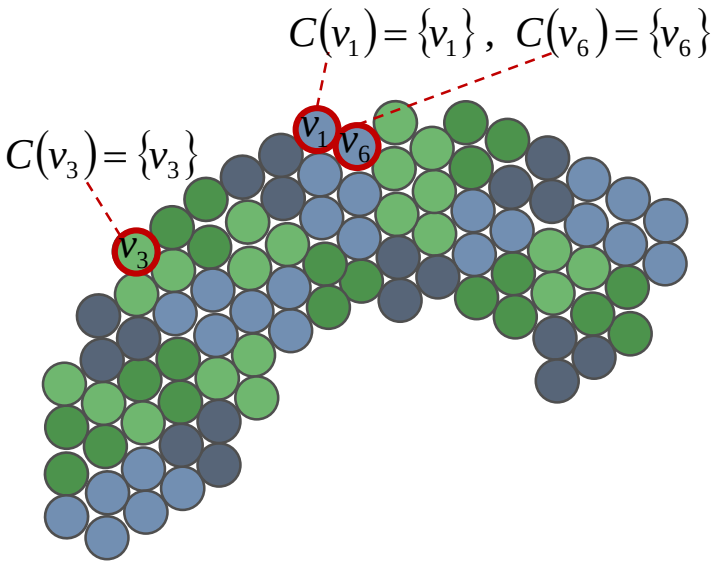


SLIC superpixels, Achanta et al, 2010

A Novel Model of Visual Context

Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

- $C(v_i)$ denotes "the coxel of patch v_i". Initially, $\forall v_i \;\; C(v_i) = \{v_i\}$.



A Novel Model of Visual Context

Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

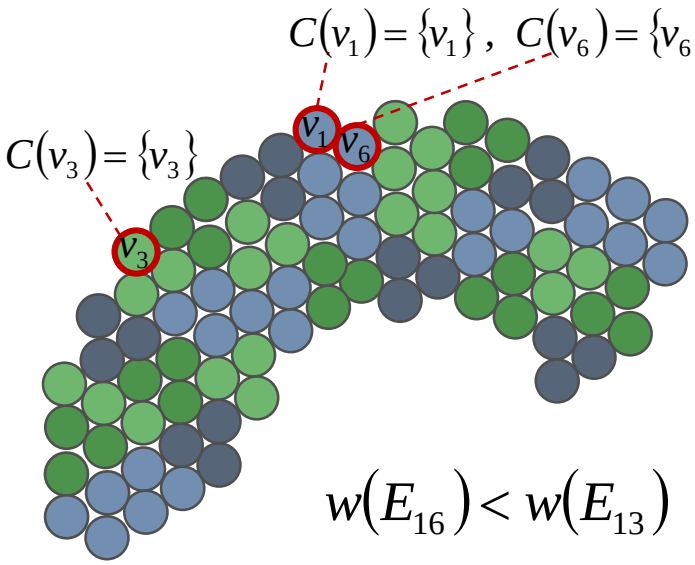
❖ The contextual gap between patches v_i and v_j:

$$w(E_{ij}) = 1 - \left(\frac{1 - \alpha \cdot D_{ij}}{1 + \beta \cdot S_{ij}} \right) \in [0,1]$$

Distance, Similarity

Lower values for closer and more similar pairs of patches. Higher otherwise.

$$0 = w_1 < \ldots < w_q < \ldots < w_r < \ldots < w_s < \ldots < w_m = 1$$



A Novel Model of Visual Context

Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

Step 1: Gradual Extension of Coxels

❖ Extension until **gap** between patches v_i and v_j:

$$If \quad w(E_{ij}) \le w_l \quad and \quad C(v_i) \neq C(v_j),$$

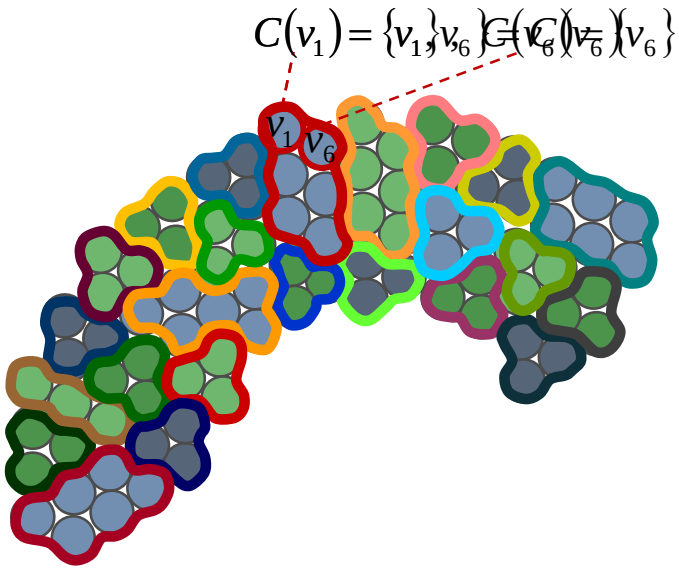
$$then \quad C(v_i) = C(v_i) \cup C(v_j)$$

$$w(E_{ij}) = \left(\frac{1 - \alpha \cdot D_{ij}}{1 + \beta \cdot S_{ij}} \right) \in [0,1]$$

Distance, Similarity

Lower values for closer and more similar pairs of patches. Higher otherwise.

$$0 = w_1 < \ldots < w_q < \ldots < w_r < \ldots < w_s < \ldots < w_m = 1$$



A Novel Model of Visual Context

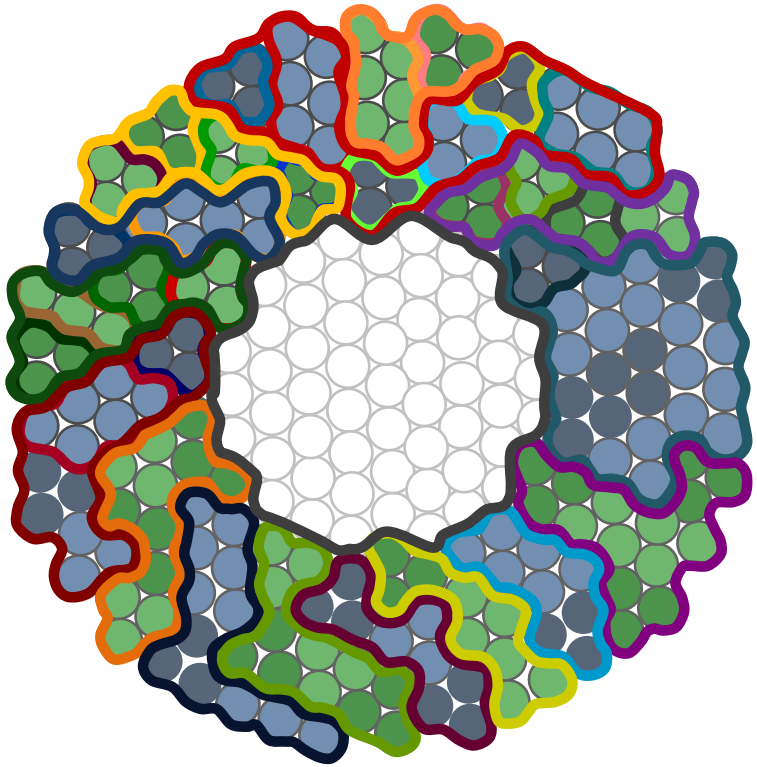
Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

Step 1: Gradual Extension of Coxels

$$0 = w_1 < \ldots < w_q < \ldots < w_r < \ldots < w_s < \ldots < w_m = 1$$

❖ **Extension rule**:

$$If \quad w(E_{ij}) \le w_l \quad and \quad C(v_i) \ne C(v_j),$$
$$then \quad C(v_i) = C(v_i) \cup C(v_j)$$



A Novel Model of Visual Context

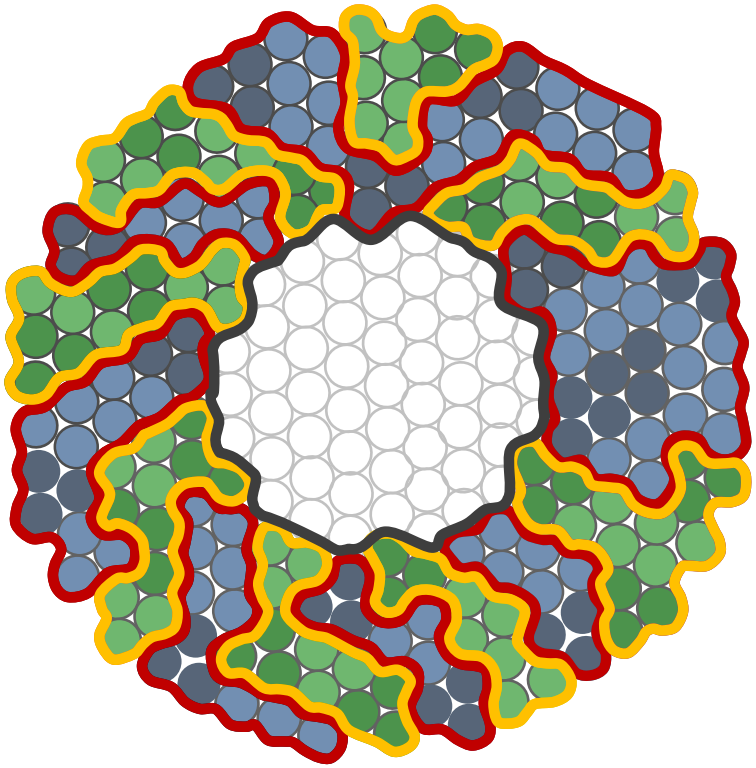
Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

Step 1: Gradual Extension of Coxels

$$0 = w_1 < \ldots < w_q < \ldots < w_r < \ldots < w_s < \ldots < w_m = 1$$

❖ **Extension rule**:

$$If \quad w(E_{ij}) \le w_l \quad and \quad C(v_i) \ne C(v_j),$$
$$then \quad C(v_i) = C(v_i) \cup C(v_j)$$



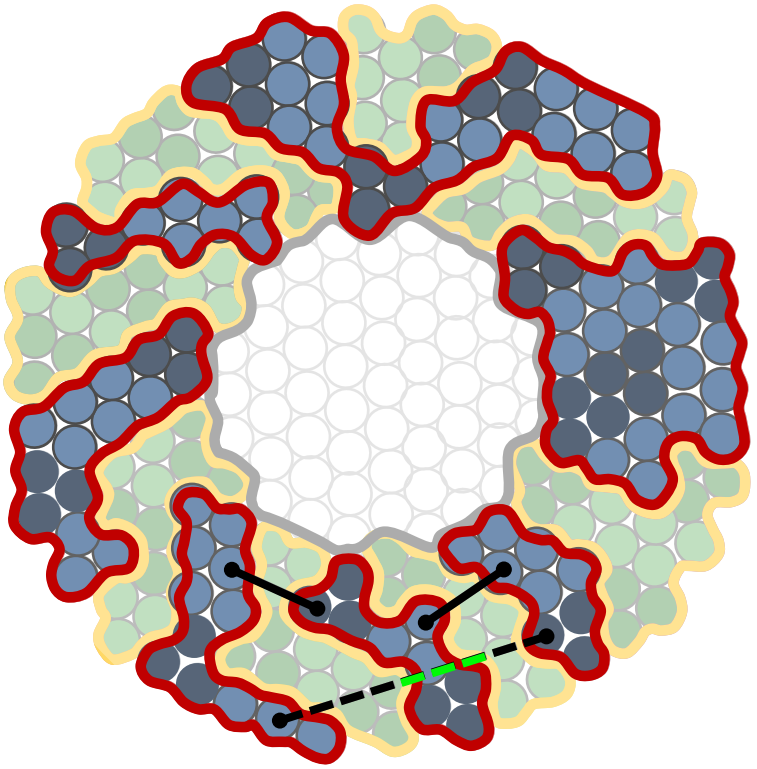
A Novel Model of Visual Context

Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

Step 2: Accumulate Saliency Votes

$$0 = w_1 < \ldots < w_q < \ldots < w_r < \ldots < w_s < \ldots < w_m = 1$$

❖ **Saliency bridge:** An edge between patches of the same coxel, that doesn't traverse another patch from that coxel.



A saliency bridge

Not a saliency bridge

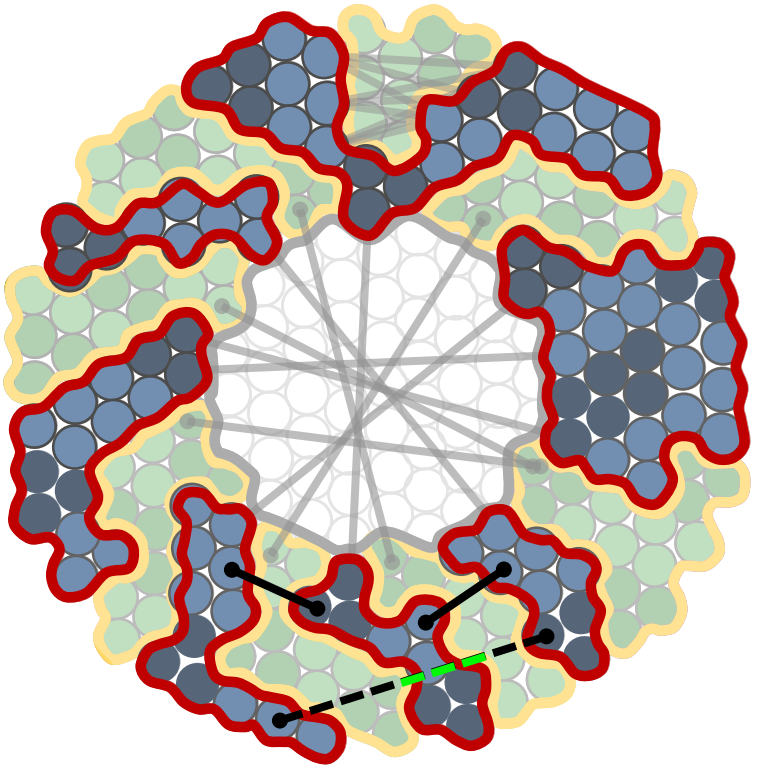
A Novel Model of Visual Context

Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

Step 2: Accumulate Saliency Votes

$$0 = w_1 < \ldots < w_q < \ldots < w_r < \ldots < w_s < \ldots < w_m = 1$$

- **Saliency bridge:** An edge between patches of the same coxel, that doesn't traverse another patch from that coxel.



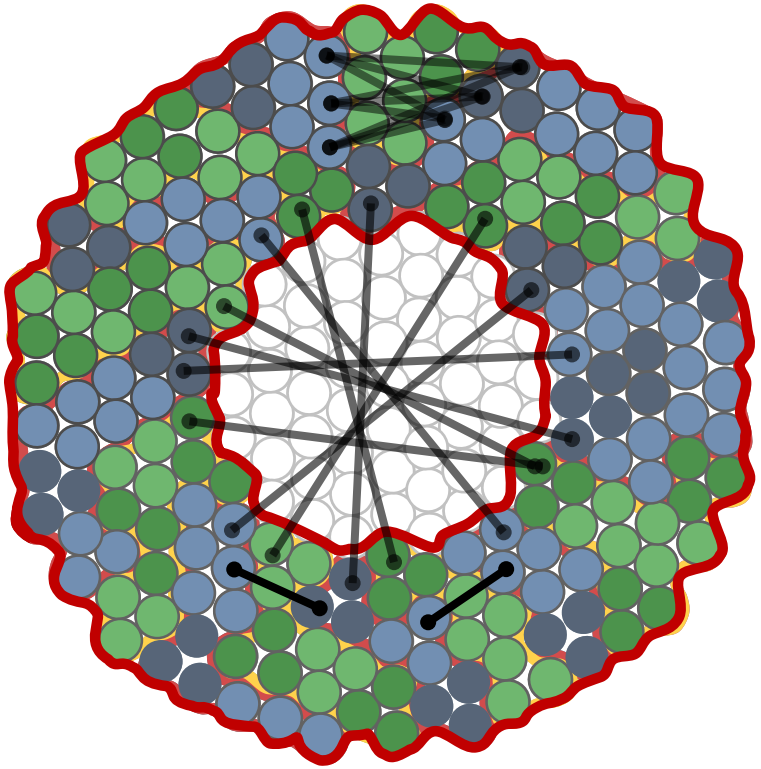
A Novel Model of Visual Context

Context Element (coxel): Smoothness, Apathy to contiguity, Enclosure.

Step 2: Accumulate Saliency Votes

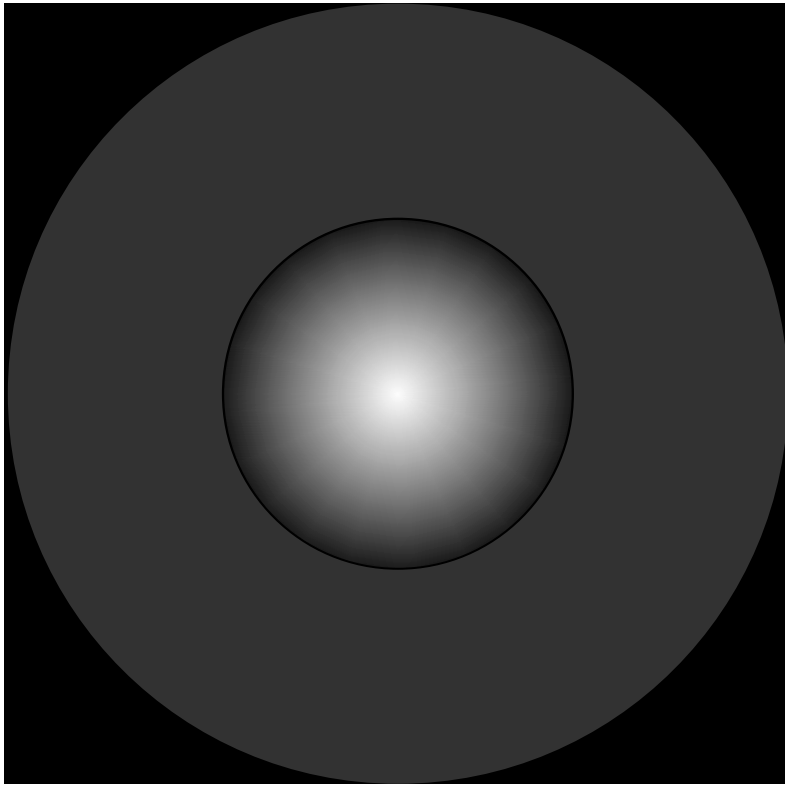
$$0 = w_1 < \ldots < w_q < \ldots < w_r < \ldots < w_s < \ldots < w_m = 1$$

- **Saliency bridge:** An edge between patches of the same coxel, that doesn't traverse another patch from that coxel.



Final saliency map

- Votes are accumulated in the image pixels saliency bridges traverse.
- A kernel density estimation is applied to obtain a dense map.



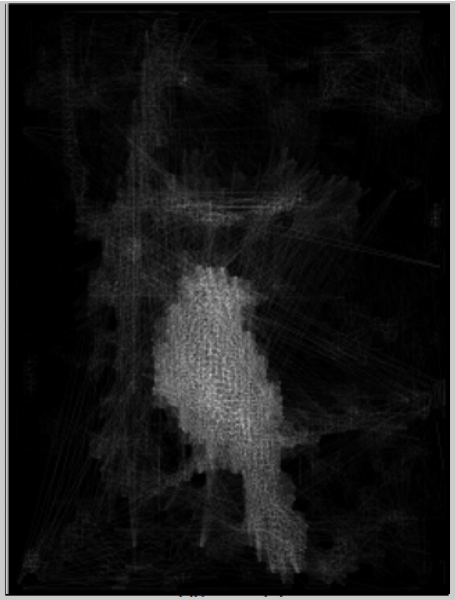
Contextual Emergence of Object Saliency

Computational process demonstrated:

Input

Saliency heat map

Final saliency map



Accumulated votes

Contextual Emergence of Object Saliency

Comparison with contrast/rarity based saliency maps:

Input

Our result

Contrast based Saliency Maps

[1] **[2]**

1. Liu et al., Learning to detect a salient object, PAMI 2011
2. Jiang et al., Automatic salient object segmentation based on context and shape prior, BMVC 2011

Contextual Emergence of Object Saliency

Comparison with contrast/rarity based saliency maps:

Input

Our result

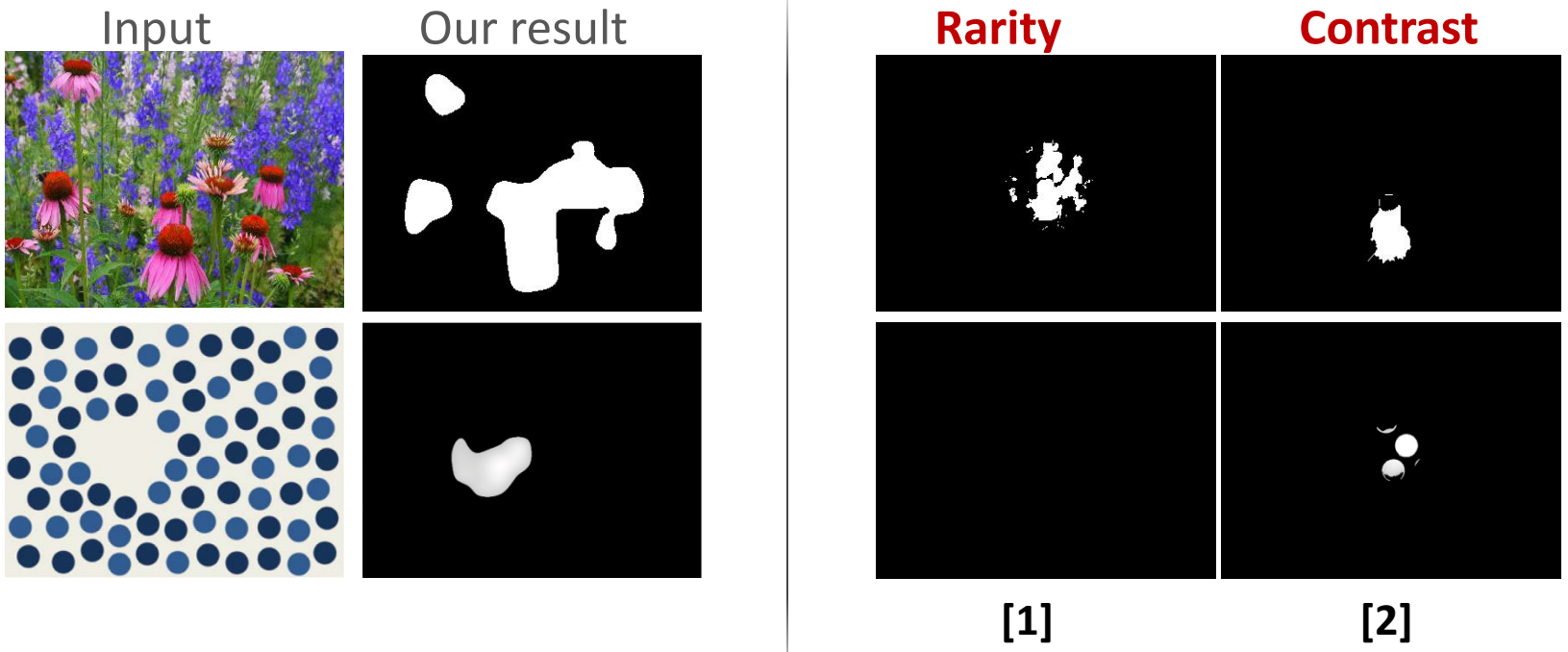
Rarity based Saliency Maps

[1] **[2]**

1. Margolin et al., What makes a patch distinct?, CVPR 2013
2. Cheng et al., Global contrast based salient region detection, CVPR 2011

Contextual Emergence of Object Saliency

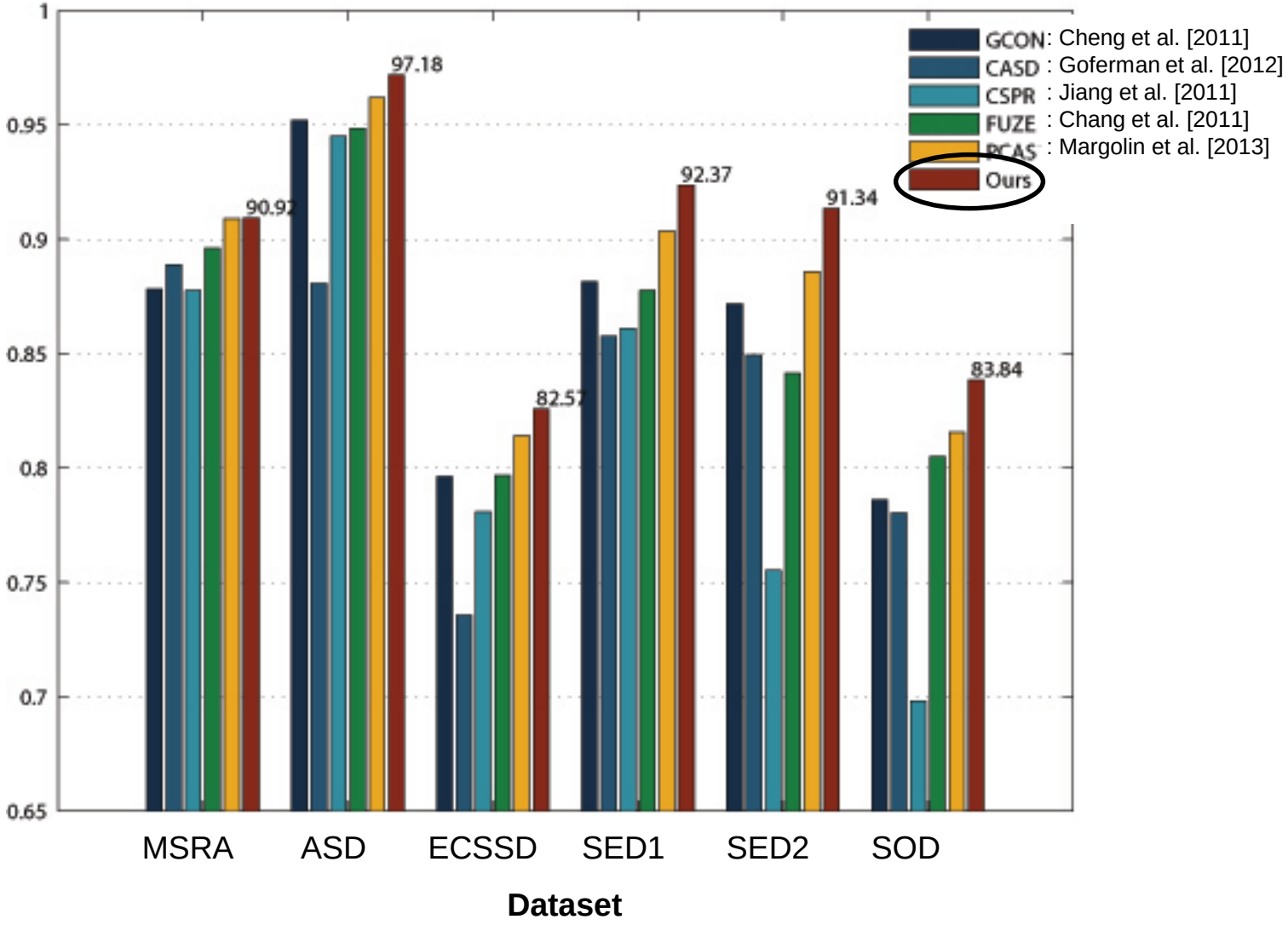
Comparison with contrast/rarity based saliency maps:



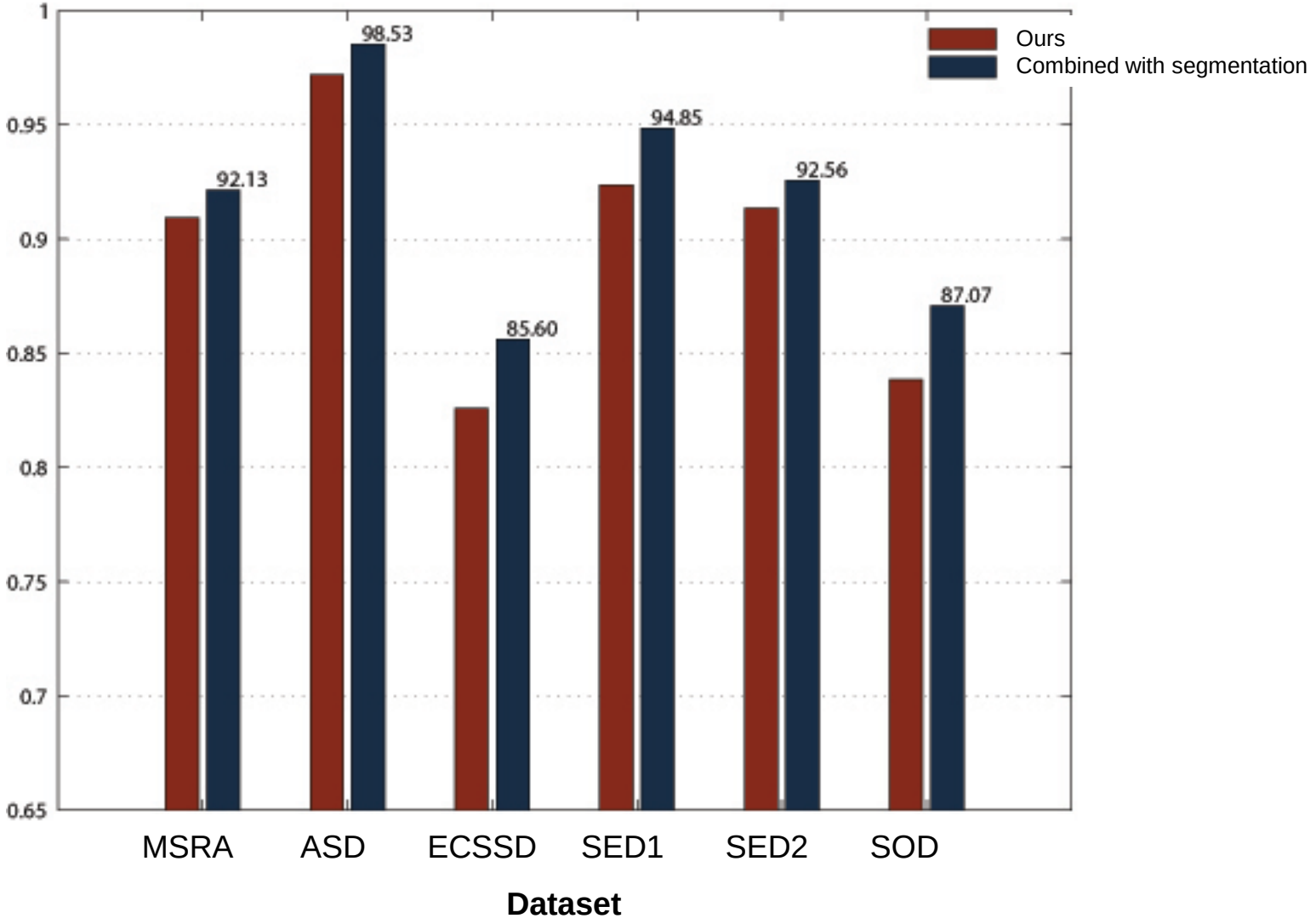
1. Margolin et al., What makes a patch distinct?, CVPR 2013
2. Jiang et al., Automatic salient object segmentation based on context and shape prior, BMVC 2011

Contextual Emergence of Object Saliency

Quantitative Evaluation:



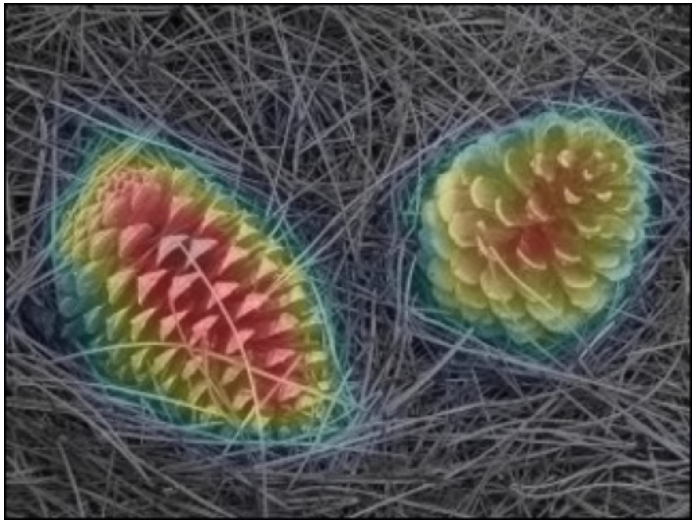
Quantitative Evaluation: Further improvement with segmentation



Contextual Emergence of Object Saliency

Summary:

- Modeling visual context may bypass the complexities (both geometrical and appearance) associated with the concept of object.
- Introduced the notion of coxels and their hierarchical construction.
- Suggested an explicit way for how saliency emerges from coxels.
- Improved state-of-the-art even before using object-related information.
- Allows to capture the saliency of complex, multiple, and abstract objects.



Thank you

European Commission SEVENTH FRAMEWORK PROGRAMME

Zlotowski Center for Neuroscience

The National Institute for Psychobiology in Israel

agriculture biological & cognitive robotics

ISRAEL SCIENCE FOUNDATION

Ben-Gurion University of the Negev