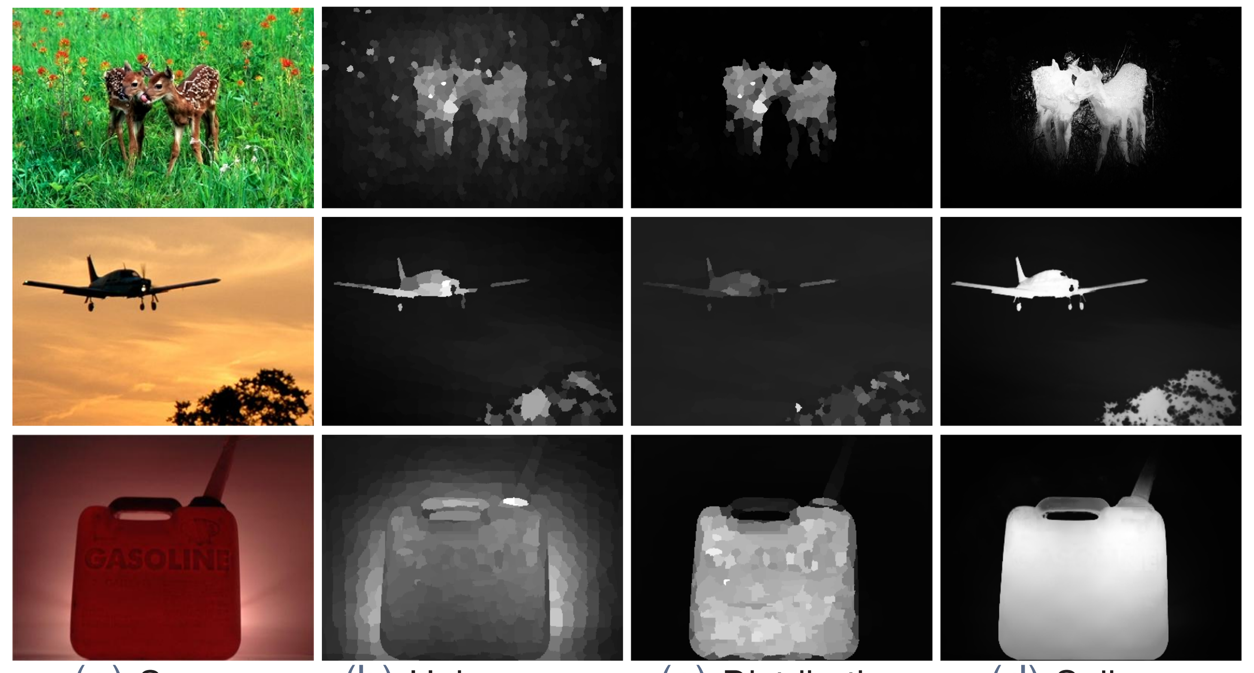
Saliency Filters: Contrast Based Filtering for Salient Region Detection

Federico Perazzi, Philipp Krähenbühl, Yael Pritch, Alexander Hornung

Disney Research Zurich, Stanford University

Abstract

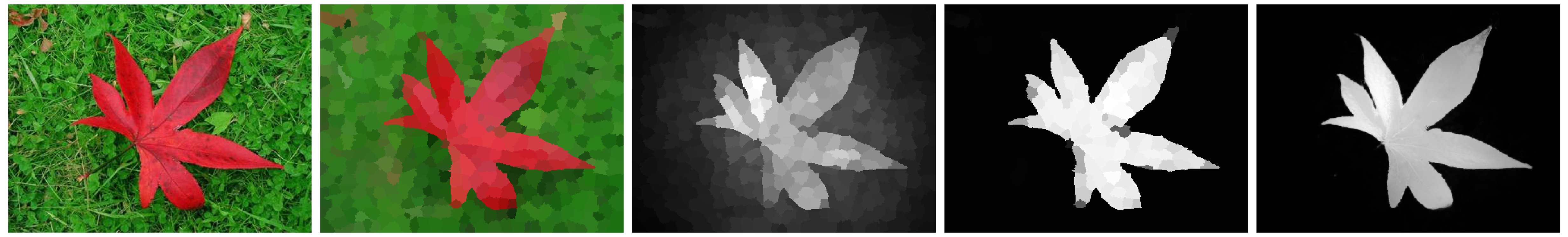


(a) Source (b) Uniqueness (c) Distribution (d) Saliency

Figure: Visual comparison uniqueness (b), spatial distribution (c), and the combined saliency map (d). It is apparent that the uniqueness measure prefers rare colors, whereas the distribution measure favors compact objects.

Saliency Estimation has become a valuable tool in image processing. In this paper we reconsider some of the design choices of previous methods and propose a conceptually clear and intuitive algorithm for contrast-based saliency estimation. Our method is based on the observation that an image can be decomposed into basic, structurally representative elements that abstract away unnecessary detail, and at the same time allow for a very clear and intuitive definition of contrast-based saliency. Our first main contribution therefore is a concept and algorithm to decompose an image into perceptually homogeneous elements and to derive a saliency estimate from two well-defined contrast measures based on the uniqueness and spatial distribution of those elements. We show that the complete contrast and saliency estimation can be formulated in a unified way using high-dimensional Gaussian filters. This contributes to the conceptual simplicity of our method and lends itself to an efficient implementation with linear complexity.

Algorithm Overview



(a) Source (b) Abstraction (c) Uniqueness (d) Distribution (e) Saliency

- **Abstraction**: Decomposes an image into compact, perceptually homogeneous elements represented by their mean color. Elements preserve relevant structure and abstract away undesirable detail.
- **Uniqueness**: Regions, which stand out from other regions in certain aspects, catch our attention and hence should be labeled more salient. Thus we measure the uniqueness/rarity U_i of each element:

$$U_i = \sum_{j=1}^{N} \|\vec{c}_i - \vec{c}_j\|^2 w_{ij}^{(p)} = \vec{c}_i^2 \underbrace{\sum_{j=1}^{N} w_{ij}^{(p)}}_{1} - 2\vec{c}_i \underbrace{\sum_{j=1}^{N} \vec{c}_j w_{ij}^{(p)}}_{\text{blur } c_j} + \underbrace{\sum_{j=1}^{N} \vec{c}_j^2 w_{ij}^{(p)}}_{\text{blur } c_j^2} . \tag{1}$$

- **Distribution**: Ideally colors belonging to the background will be distributed over the entire image exhibiting a high spatial variance, whereas foreground objects are generally more compact. Thus we measure the spatial distribution D_i of each element:

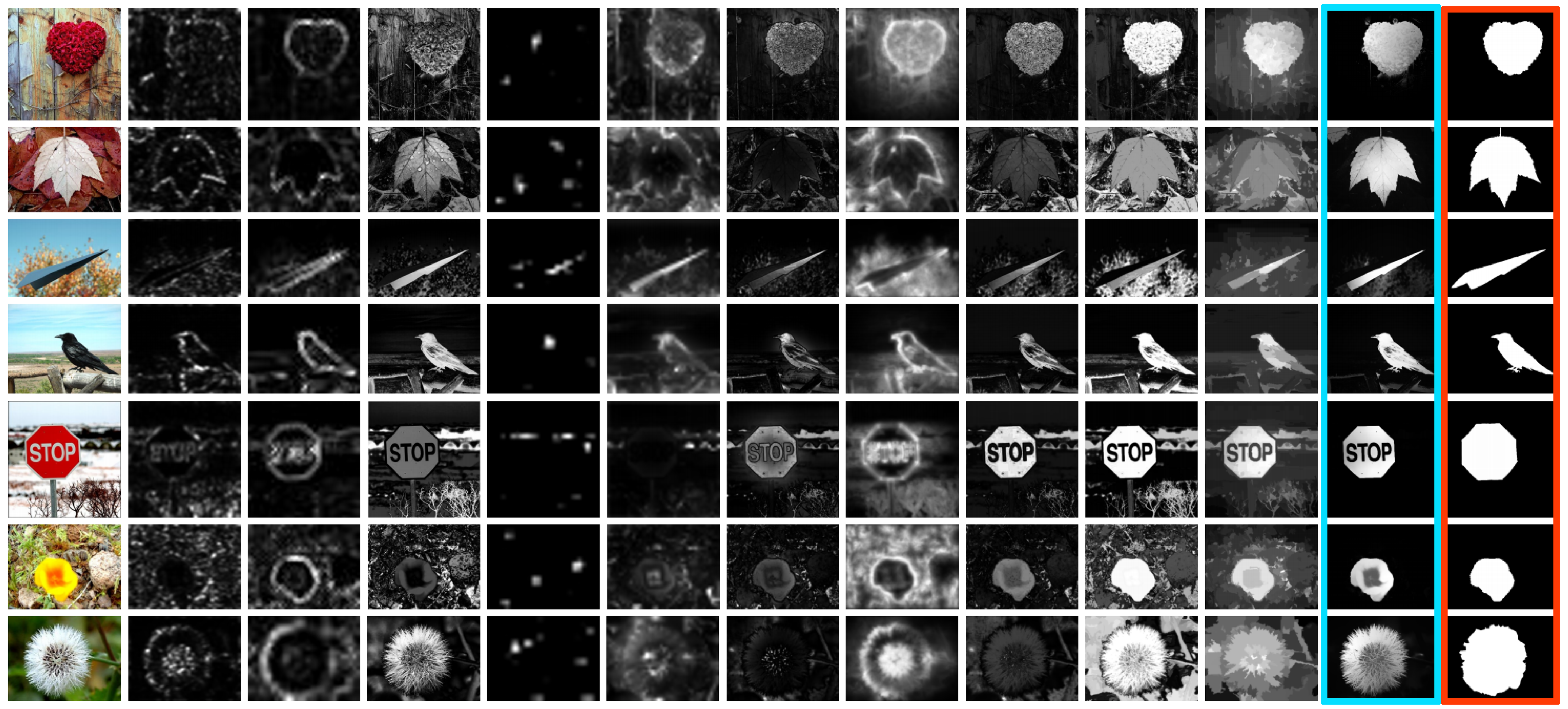
$$D_i = \sum_{j=1}^{N} \|\vec{p}_j - \vec{\mu}_i\|^2 \underbrace{w(\vec{c}_i, \vec{c}_j)}_{w_{ij}^{(c)}}, \qquad w_{ij}^{(c)} = \frac{1}{Z_i}\exp\left(-\frac{1}{2\sigma^2}\|\vec{c}_i - \vec{c}_j\|^2\right) \tag{2}$$

- **Saliency Map**: Linear combination of the saliency S_j of its surrounding image elements. S_j is the combination of uniqueness and distribution. By choosing a Gaussian weight, we ensure the up-sampling process is both local and color sensitive.

$$\tilde{S}_i = \sum_{j=1}^{N} w_{ij} S_j, \qquad S_j = U_j \cdot \exp(-k \cdot D_j), \qquad w_{ij} = \frac{1}{Z_i}\exp\left(-\frac{1}{2}(\alpha\|\vec{c}_i - \vec{c}_j\|^2 + \beta\|\vec{p}_i - \vec{p}_j\|^2)\right) \tag{3}$$

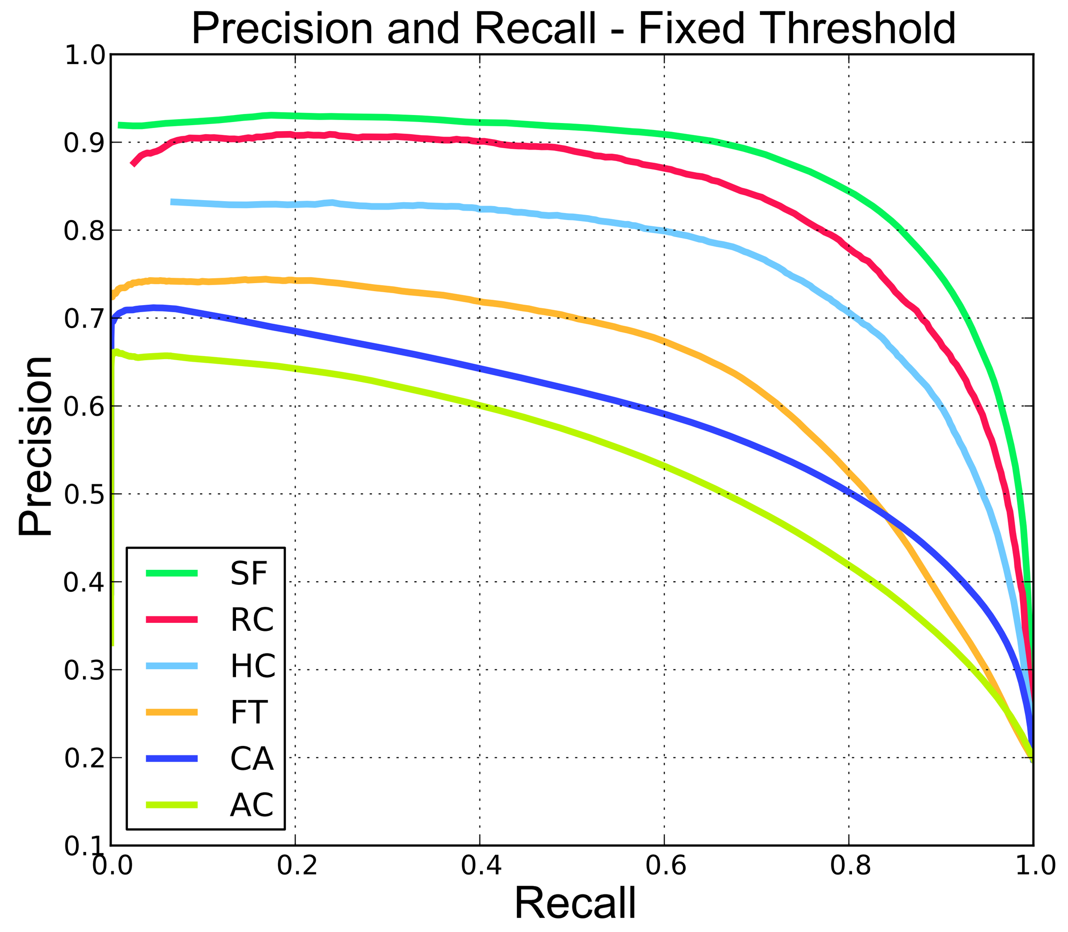
- **Implementation**: Thanks to our formulation of the above contrast measures as high-dimensional Gaussian filters, they can all be evaluated highly efficiently in the same filtering framework, using a permutohedral lattice embedding.

Results and Evaluation



(a) SRC (b) SR [16] (c) MZ [22] (d) LC [29] (e) IT [18] (f) GB [15] (g) AC [1] (h) CA [12] (i) FT [2] (j) HC [7] (k) RC [7] (l) SF (m) GT

Figure: Visual comparison of previous approaches to our method (SF) and ground truth (GT). As also shown in the numerical evaluation, SF consistently produces saliency maps closest to ground truth



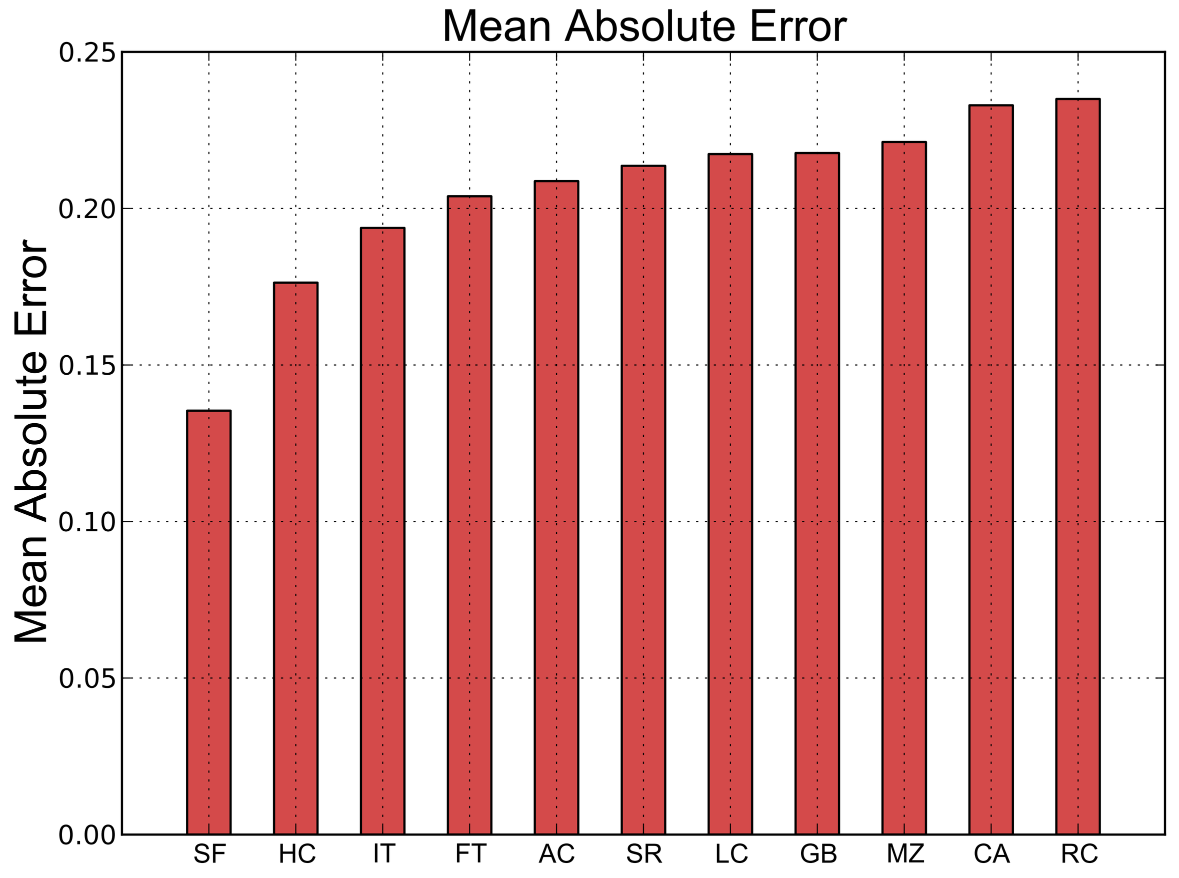


Figure: **Left**: Precision and recall rates for different state-of-the-art algorithms. **Right**: Mean absolute error of the different saliency methods to ground truth. In all experiments, our approach consistently produces results closest to ground truth