Signal Processing: Image Communication 87 (2020) 115890

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Conditional Random Field based salient proposal set generation and its application in content aware seam carving

Prerana Mukherjee [a,*], Brejesh Lall [b]

[a] Department of Computer Science, Indian Institute of Information Technology, Sri City, Andhra Pradesh, India
[b] Department of Electrical Engineering, Indian Institute of Technology, Delhi, India

ARTICLE INFO

Keywords:
Structured prediction
Edge classification
Salient edge density based scoring
Bayesian probabilistic edge map
Salient region preserving seam carving

ABSTRACT

Most of the existing generic object localization algorithms usually give the plausible object locations without taking into consideration the saliency ordering of the proposal set. This paper presents a novel object proposal generation which ranks the key objects according to their saliency score in the proposal pool. First, we formulate a Bayesian framework for generating a probabilistic edgemap which is used to assign a saliency value to the edgelets. A conditional random field is then learnt for edge-labeling by effectively combining the edge features with the relative spatial layout of the edge segments. Lastly, we propose an objectness score for the generated proposal set by analyzing the salient object edge density completely lying within the candidate boxes. Extensive experiments on the benchmark PASCAL VOC 2007 and 2012 datasets demonstrate that the proposed method provides competitive performance against popular generic object detection techniques while using fewer number of proposals. Additionally, we demonstrate the applicability of the generated proposal set for content aware image retargeting.

1. Introduction

Object localization with high degree of precision is quite a challenging task. Further, the rate of increase in the plausible set of object hypotheses is not linear to the expansion in the image size as is the case in the dense or uniform sampling paradigms for object proposal generation strategies. It is of utmost importance to generate the content aware candidate windows which would result in fewer number of bounding boxes providing good coverage of the prominent objects contained in the image. It requires significant amount of training from various exemplar models and involves meticulous selection of the object parts from potentially confusing background knowledge for machines to distinguish the precise spatial extent of the objects. To select the optimal subset of *'good'* object regions and provide a tight bound on the spatial extent of the bounding boxes involves appropriate feature selection so that the objects are characterized well. Thus, an approach to good quality object proposal generation is to leverage the strength of feature statistics. Although with the advent of deep learning based techniques [1] and the availability of huge corpus of image data the task of training a machine with huge amount of manually annotated data has eased a lot. Still, it is difficult to capture many interesting patterns like convexity and smoothness of region boundaries locally. There is scope for improvement especially in cases where a new object category appears in a scene. Therefore, the need arises for a model which captures the essence of likeliness of the object regions [2] to provide a suitable set of object proposals in a computationally efficient way alleviating the bottleneck of utilization of sophisticated resource engines.

Recently, a couple of techniques have tried to exploit the potential of edges as an object localization cue [3,4]. Edges capture most of the shape information thus preserving important structural properties contained in the image. They often occur at locations adhering to the object boundaries which make them a suitable candidate as precursor to object localization as well as segmentation. The human visual system derives important structural cues amidst chaos by perceptual organization principles to group edges to result into object-like structures. Motivated by this analogy, we construct conditional random field framework to learn the edge labeling by effectively combining the edge features with the relative spatial layout of the edge segments thus employing two key Gestalt principles [5], i.e. continuity and proximity as cues for edge grouping to facilitate structured prediction. In the proposed technique, we utilize a Bayesian formulation to construct a probabilistic edgemap in order to assign a saliency value to the edgelets by exploiting edge features which incorporate the contextual information around the edge segments. In contrast to the aligning works [3,4], the proposed object proposal generation technique takes advantage of the saliency aspect to rank the proposal set in relative order of salience while having high detection rate even with lesser number of proposals. Fewer number of high precision hypotheses reduces the number of

* Corresponding author.
E-mail addresses: prerana.m@iiits.in (P. Mukherjee), brejesh@ee.iitd.ac.in (B. Lall).

https://doi.org/10.1016/j.image.2020.115890
Received 27 April 2019; Received in revised form 28 February 2020; Accepted 21 May 2020
Available online 27 May 2020

spurious false positives [2]. The contemporary deep learning based methods provide excellent results but require huge amount of training data and sometimes initialization with *good* object hypotheses [1]. In particular, we demonstrate in the experiments that even with $\sim 10-20$ object proposals the detection rate is quite high ($53\% - 61.55\%$ at IoU = 0.5). The preliminary studies on the proposed methodology have been reported in [6]. This work is an extension of the early work to have a unified framework, detailed explanations and a comprehensive experimental evaluation for object localization. In addition to that, we show an application domain of content aware seam carving with the generated proposal set in which we pass the seams through regions of low visual importance having minimal energy in the cumulative energy map in order to prevent object distortion. The key contributions can be summarized as,

1. To the best of our knowledge, this is the first work to establish the concept of object edge classification in a Conditional Random Field (CRF) framework for object proposal generation. We further augment this knowledge with other strong features like texture and color-gradient information in the neighborhood of the edge (edge context) to assist the object proposal generation scheme.
2. We pose the edge saliency detection as a Bayesian inference problem to infer the edge segments as belonging to the object (salient) or background (non-salient). The saliency value of the edge segment is used in the objectness score of the candidate windows in order to rank key objects in relative order of salience using the generated box proposals.
3. We demonstrate the applicability of the generated proposal set for content aware image retargeting using automated seam carving approach in the images. Further, we evaluate the performance of our content aware seam carving technique over several image retargeting quality metrics exhaustively. We also compare and evaluate the performance of the proposed salient proposal generation scheme with competing object proposal generation techniques at varying IoU thresholds.

Overall architecture of the paper is as follows. Background about the related works is provided in Section 2. The proposed methodology is explained in Section 3. The content aware image retargeting via automated seam carving using the proposed object proposal set has been detailed in Section 4. Experimental results and discussion are given in Section 5 followed by conclusion in Section 6.

2. Related work

In this section we provide a detailed overview of the contemporary techniques prevalent in the domains which are closely related to our work.

2.1. Objectness prior for salient object detection

Salient object detection methods relying on various objectness measures exploit the contrast [7] or boundary priors [8]. The contrast prior helps to capture the uniqueness of the object and the background based on center–surround difference between pixels or regions. They further can be classified into local [9] and global [10] salient objects with contrast differencing methods. These constraints can be handled by class independent object proposals which directly localize the potential regions or bounding boxes where the object is present. The methods described in [2,3,11,12] result in an ordered set of proposals which contain the likely object locations.

In this paper, we evaluate a generic object proposal generation by utilizing CRF based edge classification learning framework per se, and have abstained from any form of additional supervision such as off-the-shelf saliency measures (regional saliency map computation), negative data, and pretrained features. We rank the proposals in their relative order of salience thus making the scheme suitable for applications such as adaptive compression, seam carving [13–15], scene understanding, video summarization etc.

2.2. Object hypotheses ranking

Most of the object proposal generation methods either resort to objectness cues or similarity based aggregation resulting in coherent regions. In contrast to similarity based aggregation strategies, the window generation methods provide a score to the proposal set to rank them and are considered faster [2,3,16,17]. Further, to maintain the localization accuracy the initial set of window generation is performed densely. Although some recent methods [18–20] counteract this by further refining the box bound limits. In [2], authors define object adobes in the image comprising of the salient object parts which are the potential object locations. The object adobes within the randomly sampled boxes are primarily distinguished by the local contrast difference with respect to the background. The bounding boxes are then readjusted to tightly encompass the object adobes.

We develop a saliency driven Bayesian framework to estimate edge saliency on a precomputed edge map. Furthermore, we classify the object and non-object edges using a CRF framework. We score the proposals based on the object edge density contained inside them. We incorporate the saliency value and length of the edge segments in the objectness score. This encourages the salient object proposals to be ranked higher in the proposal set thus, maintaining an inherent saliency ordering similar to adobe boxes strategy [19]. However, we do not resort to any re-ranking strategies as in [20]. Apart from this, we also demonstrate that high precision and recall rate are achieved with fewer number of proposals.

3. Proposed methodology

In this section, we present the proposed salient object proposal generation scheme. The end-to-end pipeline of the proposed method is shown in Fig. 1. As shown in Fig. 1, we first process the input image with an edge detection framework (Oriented Edge Forests) to get an initial edge map. Further, we pose the edge saliency estimation as a Bayesian inference problem to associate a saliency value to each edgelet (or edge segment) thus resulting in a probabilistic saliency map. We utilize the edge context with aggregated edge features in a Conditional Random Field (CRF) based graph-formulation to facilitate structured prediction. Finally, we calculate the objectness score of the candidate windows based on the density of salient object edges lying completely inside them. In the following subsections, we describe the components of the proposed method in detail.

3.1. Probabilistic edge saliency map computation

The primary visual cortical cells in the primate visual system responds primarily to the oriented edge responses which are then perceptually grouped to form continuous object contours. Inspired by this analogy, we devise a strategy for identifying object edge pixels and using this edge map as a strong prior for object localization. Edge detection can be done using naive gradient based techniques like Canny edge detector [21] or by sophisticated methods as described in [22]. The edgemap remains noisy even with the current state of the art techniques and hence not suitable for high level vision tasks. To this end, a sparse edge map is utilized to form a probabilistic saliency map wherein each edgelet (edge segment) is assigned a saliency value, thus providing it a distinctiveness score. In [23], the authors utilize saliency cues such as boundary connectivity and smoothness constraints in a joint optimization framework for region-level probabilistic saliency map, however we construct an edge-level probabilistic saliency map. The saliency score of each edge pixel (edgel) is computed by encoding the local edge context information i.e. texture, color gradient, edge magnitude. Since the edgels or edge pixels do not constitute any color or texture information of their own we utilize the color and texture information in the local neighborhood to incorporate the contextual information which we refer to as the *local edge context*. The edge saliency detection

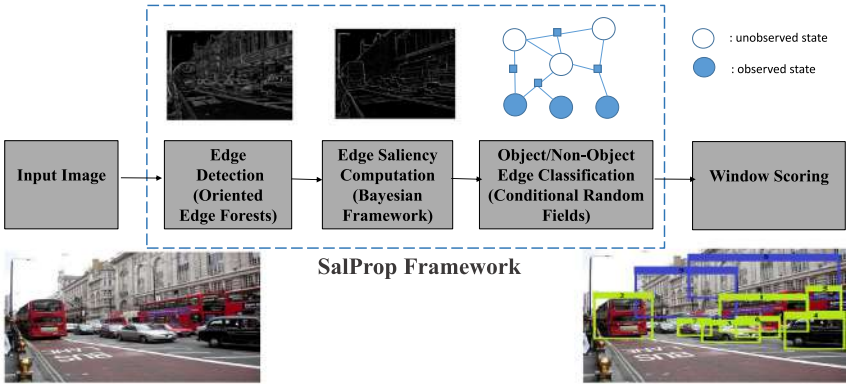

Fig. 1. The SalProp Framework. Given any RGB image, we generate proposals ranked in the order of saliency. Green boxes contain the most salient objects having higher rank and blue boxes contain less salient objects and are ranked lower in the proposal set. The number assigned to each box indicates its saliency ranking in the proposal pool. The states in the CRF framework denote the status of the node label in the graph.

is posed as a Bayesian inference problem to infer the likelihood of the edge segments as being object (salient) or background (non-salient) edge segments. In this context, since the background edge segments belong to low magnitude spurious edges or highly textured regions they do not represent the object contours and are considered as non-salient. The reason behind the choice of a Bayesian framework is the requirement of association of a saliency value with the edge segment based on the prior distribution of edge magnitudes. We estimate the prior distribution of salient or background edges based on their edge magnitude since stronger edges are more likely to be a part of an object. Apart from relying on the Bayesian inference from the edge strength, we also consider edge context to make the inference more reliable and accurate. The first step involves the computation of edge responses using Oriented Edge Forests (OEF) boundary detector [24] which is highly efficient in detecting object boundaries. In order to handle the variations due to fine clutter and texture from surface roughness and lighting conditions manifested in the local gradients it is prudent to learn boundary representations from the training data. OEF consists of random forest decision classifiers that analyze the local patches and gives the probability distribution over the entire space of oriented edges that lie in that patch as the output. Finally, the predictions over the local patches are calibrated and fused over the image pyramid to obtain a final oriented boundary map. Due to the structured prediction of the boundary contrast at each hypothesized edge orientation it produces a more reliable boundary representation with less memory footprint for training the model, high computational speedup and accuracy as compared to traditional methods as in [22,25]. We utilize the sparse variant of OEF detection in which non-maximal suppression (NMS) is used. The resultant sparse edge map consists of each pixel i having an edge magnitude $|e_i|$. Next, we further perform a thresholding (provides computational efficiency) by only considering edge segments with length $l > \alpha$ and edge pixels having magnitude $|e_i| > \beta$.

Given an edge segment having a relative edge strength denoted as $\mathfrak{s}$, the posterior probability $p(sal|\mathfrak{s})$ of the edge segment in the sparse edge map is mathematically formalized as:

$$p(sal|\mathfrak{s}) = \frac{p(sal)p(\mathfrak{s}|sal)}{p(sal)p(\mathfrak{s}|sal) + p(bg)p(\mathfrak{s}|bg)}, \tag{1}$$

where $p(sal|\mathfrak{s})$ is the probability of the edge segment being salient. $p(sal)$ and $p(bg)$ are the prior probabilities of the edge segment being salient (object edges) or background respectively. $p(\mathfrak{s}|sal)$ and $p(\mathfrak{s}|bg)$ are the likelihood of observations. Here, observation outcome is the probability of the edge segment being salient or background given the set of parameter values i.e. edge features. Edge saliency prior of an edge segment is computed as:

$$p(sal) = \frac{\mathfrak{N}}{max_j \mathfrak{N}_j}, \mathfrak{N} = f_G.f_{LTP}.\mathfrak{s}, \tag{2}$$

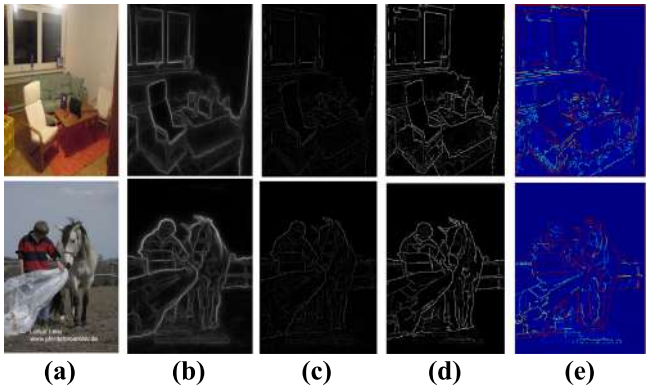

Fig. 2. (a) Original image (b) Edge map using OEF (c) After NMS and thresholding (d) Bayesian Probabilistic edge map (indicating saliency of edge segments) (e) Colormap of Bayesian Probabilistic edge map.

where $\mathfrak{N}$ indicates the scalar multiplication of the texture (f_{LTP}), color (f_G) and edge magnitude values of the edge pixels in the edge segment ($\mathfrak{s}$) (as given in Section 3.1.1), j is an index over edge segments. We utilize the normalized edge saliency prior value for an edge segment.

The background prior is given as,

$$p(bg) = 1 - p(sal). \tag{3}$$

To find the likelihood, we need to separate the edge segments into salient or background segments. If the edge magnitude $\geq \beta_1.\mathfrak{s}$, we consider it as salient, else it is considered to be a background edge segment. Here, β_1 indicates the edge magnitude threshold, where $\beta_1 > 0$. We then compute the normalized histograms h_s and h_{bg} of the edge magnitudes of the edge pixels in salient and background edge segments respectively with 10 bins each. The observation likelihoods $p(\mathfrak{s}|sal)$ and $p(\mathfrak{s}|bg)$ are calculated from h_s and h_{bg} respectively based on bin value to which $\mathfrak{s}$ of the edge segment belongs. The probabilistic edge map is shown in Fig. 2. As can be observed in the sample images, the OEF output appears to be smudgy so to have a single pixel thickness edge map we perform some post-processing steps (thresholding and NMS operations) on the OEF edgemap. NMS helps in suppression of the points that do not lie in the important edges and thus produces edge peaks. The salient (object) edges are more accentuated than the background (non-object) edges in the probabilistic edge map. The probabilistic edge map assists in ranking the candidate windows by assigning an objectness score to them (as explained in Section 3.3).

3.1.1. Edge features used in Bayesian edge saliency formulation

In order to quantify the directional change in the intensity or color in an image we compute the color gradients along the edge

segments along four dominant orientations-horizontal edges, vertical edges and diagonally oriented edges. It is computed by convolving the Gaussian kernel of a particular orientation with the image. We integrate the magnitudes of color gradients of a particular orientation ($G_{o,i}$), $o\epsilon\{0°, 45°, 90°, 135°\}$ along the edges denoted by f_G, given as:

$$f_G = \sqrt{\sum_o (\sum_i G_{o,i})^2}. \tag{4}$$

The chosen kernel is a weighted Gaussian filter to handle noise variations. With increase in kernel size there is an increase in smoothness and thus loss of texture variations. So, we have chosen a 3x3 filter which retains such perceptual information. Fig. 3 shows the results of the color gradient on the image at different dominant orientations.

Local Binary Pattern (LBP) features are known to be sensitive to noise in uniform image regions as they are based on crisp thresholding. LBP is invariant to monotonic intensity changes, thus it is robust to illumination and contrast changes. However, it is sensitive to noise and small pixel value fluctuations. Therefore, Local Ternary Patterns (LTP) [26] can handle this situation as it is more discriminant and less sensitive to noise in uniform regions. We utilize a variant of LTP [26] to capture the textural variations in the local edge neighborhood. f_{LTP} is the local ternary pattern (LTP) of the edge pixels I_i contained in the given edge segment. It is computed by comparing the intensity value of the current pixel with the intensity values of its neighbors denoted by I_{nb} using a kernel of size 3x3. A 3^B-bin block histogram is computed in case of LTP codes. For $B = 8$, the histogram has 6561 bins which is high-dimensional. However, in [26], the authors utilize the LTP code as a combination of its "upper" and "lower" local binary pattern (LBP) codes. Due to this operation, the bin number is reduced from 6561 to 512. Since, we represent LTP for the edge segments only we take the average variance of this combination over the edge. Here, T is a user defined threshold and B is the total number of neighboring pixels. We compute the variance of all the LTP values of the edge pixels for a particular segment given as:

$$f_{LTP} = \frac{\sigma(ULBP) + \sigma(LLBP)}{2}, \tag{5}$$

$$ULBP = \sum_{b=0}^{B-1} s'(I_{nb} - I_i).2^b, \tag{6}$$

$$LLBP = \sum_{b=0}^{B-1} f'(I_{nb} - I_i).2^b, \tag{7}$$

$$s'(z) = \begin{cases} 1 & z \geq T \\ 0 & otherwise \end{cases} \quad f'(z) = \begin{cases} 1 & z \leq -T \\ 0 & otherwise \end{cases} \tag{8}$$

The maximum magnitude value $\mathfrak{s}$ (relative edge strength) of edge pixels in the edge segment is computed as follows:

$$\mathfrak{s} = max_i(|e_i|). \tag{9}$$

3.2. CRF framework for object edge classification

We formulate Edge Feature Graph Conditional Random Field (CRF) model to learn the conditional distribution over the edge segment labeling using the local edge context. CRF is used here for structured prediction for the edge labeling problem. To incorporate two key Gestalt principles, i.e. (i) continuity and (ii) proximity as cues for edge labeling, the edge segments (nodes) which are spatially close and fall in the same contour are linked in the graph. The nodes are associated with 7-D feature vector (Section 3.2.1). CRF training details have been provided in Section 5.1. The score associated with each link is denoted as e_{ij} given by a 4-D feature vector [Up/Down, Right/Left, mean, variance]. The first two elements (0/1) in the vector denote the relative position of node i with respect to node j. The next two elements denote the mean and variance in the feature differences between the two nodes in the graph. CRF framework learns the edge labeling inference by effectively combining the edge features with the relative spatial layout of the edge segments thus employing perceptual organization principles to facilitate structured prediction. The objective function (energy) of the structured prediction is given as:

$$E(L|X) = \sum_{i\epsilon\mathcal{V}} \boldsymbol{\phi}(l_i, X; \mathbf{W_1}) + \sum_{\{i,j\}\epsilon\mathcal{E}} \boldsymbol{\psi}(l_i, l_j, X; \mathbf{W_2}), \tag{10}$$

where L is the structured label, X is the structured input features, l_i is the label of the node, $\mathbf{W_1}$ are the node parameters, $\mathbf{W_2}$ are the link parameters, $\phi(l_i, X; \mathbf{W_1})$ are unary potentials given as the inner product of the node features with node weights and $\psi(l_i, l_j, X; \mathbf{W_2})$ indicates pairwise potentials given as a linear function of link features and weights (shared over all links). The objective function is optimized using Block-coordinate Frank Wolfe Structured SVM to compute $\mathbf{W} = [\mathbf{W_1} \quad \mathbf{W_2}]$. Algorithm 1 details the CRF structured prediction.

Algorithm 1 Algorithm for CRF learning and Prediction

procedure CRF–STRUCTURED PREDICTION
Input: $\mathcal{V}$: set of edge segments with 7-D feature vector, $\mathcal{E}$: set of edge links with 4-D feature vector
Output: Labels L for each node (edge segment)
Optimize an objective function (energy) with respect to parameter vector $\mathbf{W} = [\mathbf{W_1} \quad \mathbf{W_2}]$
$E(L|X) = \sum_{i\epsilon\mathcal{V}} \phi(l_i, X; \mathbf{W_1}) + \sum_{\{i,j\}\epsilon\mathcal{E}} \psi(l_i, l_j, X; \mathbf{W_2})$
where L is the structured label, X is the structured input features, l_i is the label of the node, $\boldsymbol{\phi}(l_i, X; \mathbf{W_1})$ are unary potentials and $\boldsymbol{\psi}(l_i, l_j, X; \mathbf{W_2})$ are the pairwise potentials.
Encode the structure of the problem in a joint feature function $\hat{\psi}(\mathbf{x}, \mathbf{y})$ as in prediction using,
$\hat{\mathbf{y}} = argmax_{\mathbf{y}\epsilon Y} \mathbf{W}^T \hat{\psi}(\mathbf{x}, \mathbf{y})$
where $\mathbf{y}$ is the structured label, $\mathbf{x}$ is the feature vector of a data point (node), Y denotes set of all possible labels {0,1} and $\hat{\mathbf{y}}$ is the prediction of the data point. Solve for $\hat{\mathbf{y}}$. Once the weights are learnt using the training data, predictions are made using Alternating Directions Dual Decomposition (AD3) inference algorithm.
end procedure

3.2.1. Local edge features used in CRF framework

We consider two image patches with radius r_d on either side of the edge segment to take into account the contextual information around the edge. We uniformly sample half of the data points (pixels) in region A_1 (upper local patch) and A_2 (lower local patch) to avoid overfitting. Next, we compute the textural features for the data points in these regions. For this, we compute a 5-dimensional filter bank at scale k_1. We use perceptually uniform CIELab color space. The filter bank consists of Difference of Gaussian (DoG) at 2 scales $\{k_1, 2k_1\}$, Laplacian of Gaussian (LoG) at 3 scales $\{k_1, 2k_1, 4k_1\}$. These filters are applied only to the luminance channel in order to preserve the original colors in the image and quantify only the brightness contrast. Then, we compute the variance of the DoG and LoG feature vectors of each region. We concatenate the feature vectors in ascending order of variance as $[DoG_1, DoG_2]$ and $[LoG_1, LoG_2]$. The intuition behind this is that the region having low texture variation is likely to belong to object region and vice versa while maintaining an ordering for CRF training. Thus, the 7-D feature vector for each edge segment is represented by the vector, $[f_G, DoG_1, DoG_2, LoG_1, LoG_2, f_{LTP}, \mathfrak{s}]$. We again utilize here the same features as used in edge saliency computation namely color gradient f_G, local ternary pattern f_{LTP} and relative edge strength $\mathfrak{s}$ as been explained in Section 3.1.1. To demonstrate that the 7-D feature vector represent independent characteristics, we plot the correlation matrix between them in Fig. 4. It is observed in general here that the chosen feature set has minimum correlation between them thus encoding the edge strength, texture and intensity variations similar to the center–surround difference operator around the neighborhood of the edge pixels.

3.3. Window generation and scoring

We proceed with a sliding window technique for proposal generation over position, scale (σ) and aspect-ratio (τ). Each successive window maintains an Intersection over Union (IoU) with the previous window and the step size is calculated accordingly. The IoU metric is taken as 0.65 (as in EdgeBoxes approach [3]. Scale is set from 0.5% to 95% of the image size with 1% increment between scales. The aspect

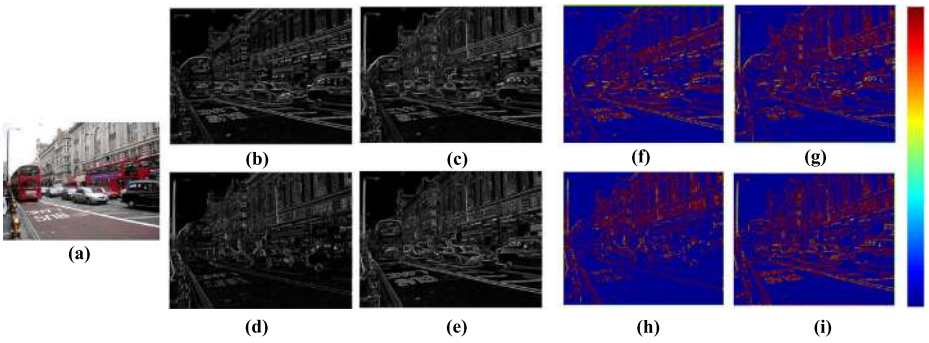

Fig. 3. (a) Original image (b)–(e) Color gradients at 0°, 45°, 90° and 135° respectively (f)–(i) Corresponding Colormap visualization (in jet colormap). Blue color corresponds to lowest edge intensities and Red color corresponds to highest edge intensities.

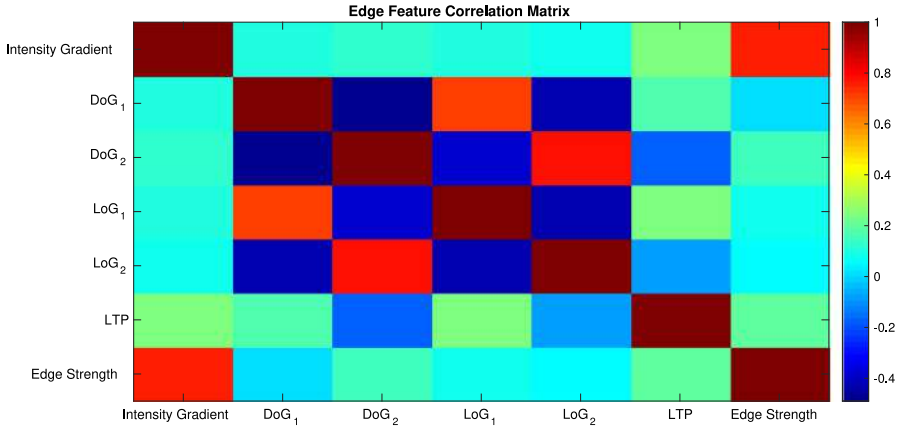

Fig. 4. Correlation matrix for the edge segment features [f_G (intensity gradient), DoG_1, DoG_2, LoG_1, LoG_2, f_{LTP}, $\mathfrak{s}$ (edge strength)]. Red color corresponds to higher correlation between the features.

ratio ranges from 1/3 to 3. The horizontal (S_h) and vertical (S_W) strides for generating bounding boxes are computed as:

$$S_H = w\left(\frac{1-\delta}{1+\delta}\right), S_W = h\left(\frac{1-\delta}{1+\delta}\right), \tag{11}$$

where δ is the required IoU between two consecutive windows, h and w are the height and width of the window respectively. $h = \sqrt{\frac{\sigma.imgArea}{\tau}}$ and $w = \sqrt{\sigma.imgArea.\tau}$ for successive candidate windows, where σ and τ denotes the scale and aspect-ratio of the window. $imgArea$ denotes the area of the image.

All edge segments that fall completely inside the proposal window increase the score depending on their edge length and saliency value. Since, we associate the saliency value of the edge segment in the objectness score we are able to rank the proposal set in their relative order of saliency. The ranking framework does not incorporate any form of additional supervision in the form of negative hard mining of data, pretrained feature set or explicit computation of regional saliency map computation which leads to additional overhead. The computational efficiency of a Bayesian probabilistic saliency map based on edge features expedites the process of saliency score association to the candidate windows. Furthermore, the objectness score discourages larger windows to have high scores by dividing the score by the area of the window given as,

$$S_w = \frac{\sum_j s_j.l_j}{\sqrt{Area_w}}, \tag{12}$$

where s_j indicates the saliency value, l_j is the length of the jth edge segment and $Area_w$ is the area of window w. There are two necessary post processing steps for generating better proposals: Refinement and Non-Maximal Suppression (NMS). We perform these steps in congruent lines to those in [3].

4. Application: Content aware seam carving

Content aware image resizing finds numerous applications in real life such as rendering web pages over smartphones or tablets, photo editing tools. Cropping or scaling may be used to effectively resize as per requirement, but they result in information loss in the image, and also impact its quality (in case of scaling). Even the traditional seam carving methods [27,28] fail to prevent image distortion in certain cases. This facilitates the need to improve the existing techniques for adaptive image retargeting by content aware resizing.

To address the issues mentioned above, we propose an improved seam carving technique that results in minimizing the perceptual loss. For this, we make use of SalProp object proposals which ranks the key objects in an image. We then apply Region of Interest (ROI) based guided filtering to preserve the boundaries adhering to the objects within these object proposals. We enhance the energy of the edge pixels lying in these object proposals using guided filtering approach so that the seams trail the path of minimum energy preserving object boundaries. We first create an energy map for the image using Canny filter. The energy of each pixel for an image I is measured by an energy function which is computed as,

$$E(I) = \left|\frac{\partial I}{\partial x}\right| + \left|\frac{\partial I}{\partial y}\right|. \tag{13}$$

A seam is a connected path of lowest energy pixels in the image, which blend effectively with their surroundings and removing which will not pose the threat of significant information loss in the image. These paths can be vertical or horizontal or both.

After the energy map computation, we find the cumulative energy map of the image by employing the obtained energy map using Dynamic Programming given as,

$$C(i,j) = E(i,j) + min[C(i-1,j-1), C(i-1,j), C(i-1,j+1)], \tag{14}$$

where $C(i,j)$ denote cumulative energy of a pixel i,j and $E(i,j)$ corresponds to the energy of the pixel at (i,j) as obtained from the energy map.

Algorithm 2 Algorithm for finding the seam with lowest energy

```
procedure FINDSEAMWITHLOWESTENERGY(x)
    Input: C: Cumulative Energy Map of the image.
    Output: SeamVector containing the column nos. of pixels to be removed.
    [rows, cols] ← size(C)
    for i ← rows to i ← 1
    if i == rows
    index ← min(C(rows,:))
    elseif SeamVector(i+1)== 1
    index ← min[∞, C(i, SeamVector(i+1)), C(i, SeamVector(i+1)+1)]
      elseif SeamVector == cols
    index ← min[C(i, SeamVector(i+1)-1), C(i, SeamVector(i+1)), ∞]
    else
    index ← min[C(i, SeamVector(i+1)-1), C(i, SeamVector(i+1)), C(i, SeamVector(i+1)+1)]
    end
    SeamVector(i)← index
    return SeamVector
    end
end procedure
```

Once the cumulative energy map has been constructed, we can start forming seams from the last row (as in Algorithm 2).

First the pixel with the least cumulative energy is identified and its index is stored in $SeamVector$. The energy map is again calculated after each seam removal, as the pixels would have their importance values altered, even if slightly. After that from the remaining pixels least energy ones form the second seam, and so on. The stopping criteria for the iterative process is when the desirable number of seams are removed from the image (until the desired dimensions are reached). In case of expansion, instead of removing them, requisite number of seams alongside the formed seams are placed.

5. Experimental results

5.1. Evaluation setup and datasets

We utilize Pystruct 0.2.5 structured prediction [29] for implementing CRF model. The CRF model is trained on the MSRA1000 saliency dataset [30] which has been chosen due to higher distinction of edge features between the object and background. The CRF training is performed in two steps. First, the edgemap is extracted using OEF followed by NMS and thresholding. We utilize the pretrained model of OEF boundary detection algorithm [24] for generating the prior boundary map. Next, we perform k-means clustering on edge magnitude of edges (with $k = 2$) to segregate them into object and non-object edges. We take the ground truth edges and higher magnitude edges as object edges while lower magnitude edges as non-object edges.

CRF is trained by utilizing the edge features as discussed in Section 3.1.1. The model is further evaluated on PASCAL VOC 2007 and 2012 datasets [35]. The parameter setting used in Section 3.1 involves $\alpha = 15$, $\beta = 40$, $T = 5$, $B = 8$, $\beta_1 = 0.8$ and in Section 3.1.1 parameter set as $k_1 = 0.5$, $r_d = 5$. These values provided best results in our experimental analysis.

We have evaluated our seam carving results on the MSRA1000 saliency dataset [30]. It consists of 1000 images with salient objects present in them. We report the results on a saliency dataset as we show the efficacy of the content aware seam carving method which is guided by a salient ordered proposal set. We utilize Top-10 object proposals only. We have compared the proposed Content Aware Seam Carving (CASC) method for 25% and 50% image reduction/carving with the following methods: (i) Cropping (CR), (ii) Scaling (SCL), (iii) Seam Carving (SC) by Avidan et al. [27], (iv) Seam Carving (SCMIT) by Rubinstein et al. [28], (v) Saliency map guided seam carving using Graph based visual saliency (SC-GBVS) [28,36], (vi) Discriminative Regional Feature Integration (SC-DRFI) [28,37], (vii) DeepSaliency (SC-DEEPSAL) [28,38] (viii) Scale and Stretch (SNS) [39], (ix) Prior-model Guided Depth-enhanced Network (SC-PDnet) [28,40], (x) modified Discriminative Regional Feature Integration (SC-mDRFI) [28,41], (xi) CASC-EdgeBoxes [3] and (xii) CASC-MCG [12]. We conduct two set of experiments to highlight the effectiveness of the proposed method in content aware seam carving. In the first set of experiments, we compare the proposed method with other state of the art seam carving and image retargeting methods using Image Retargeting Quality Assessment (IRQA) metrics such as SIFT flow [31], Earth Mover's Distance (EMD) [32] and Aspect Ratio Similarity (ARS) [33]. In the second set of experiments, we determine the distortion rates specific to the information loss index in the saliency aspect using the performance metrics such as Bi-Directional Salient Information Loss measurement (BDSIL) [34] (which consists of, Forward Saliency Information Loss (FSIL) and Backward Saliency Information Loss (BSIL)) and Salient Global Structure Distortion Measurement (SGSD) [34]. Table 1 details the relevance of these performance metrics used.

5.2. Quantitative evaluation

Validation Testing. We experiment with various variants of our approach on the validation set Pascal VOC 2007 2510 images. Fig. 5(a) shows cut-off NMS threshold for varying IoU values when generating 1000 object proposals. The cut-off value of NMS threshold is the value at which highest recall rate is achieved. At IoU = 0.5, NMS = 0.5; IoU = 0.6, NMS = 0.6 and IoU = 0.7, NMS = 0.7. These NMS thresholds serve as the values for the three variants of the proposed SalProp algorithm during testing. Three useful variants are SalProp5, SalProp6, and SalProp7 that have settings for NMS threshold adjusted for IoU thresholds $\delta = 0.5$, 0.6 and 0.7 respectively. In Fig. 5(b), we show the SalProp algorithm variant (SalProp7) giving the detection accuracy with and without center-box edge removal and bounding box refinement strategy. It has been postulated in [2] that the edges in the center of the box are of less relevance than the edges near the bounding box boundary. To handle this, the edge magnitudes of the inner box inside the candidate bounding box is subtracted from the final objectness score of the candidate bounding box. The sum of the edge magnitudes in the inner box is computed using an integral image. The inner box is of dimensionality $h/2$ and $w/2$ where h and w is the height and width of the candidate bounding box. However, in our experiments we observe that there is no performance gain on center-box removal strategy. In the refinement approach, all bounding boxes generated by the sliding window search with a score above a small threshold are refined similar to that in EdgeBoxes technique [3]. This gives a performance boost of 0.5% in detection accuracy at Top-1 proposal (see Fig. 7).

We compare the performance of SalProp algorithm against competing object proposal generation algorithms: (a) Geodesic Object Proposals [46] (b) Objectness [2] (c) Rahtu [42] (d) Randomized Prim's [43] (e) Selective Search [11] (f) Perceptual Edge [4] (g) Multiscale Combinatorial Grouping [12] (h) Rigor [45] (i) EdgeBoxes70 [3] (j) Rantalankila [44] and show that the proposed technique achieves competitive performance against these methods while using fewer number of proposals. Table 2 compares SalProp against other competing algorithms. Fig. 6(a)–(c) shows the detection rates when we are varying the number of object proposals at different IoUs and Fig. 6(d)–(f) shows the detection rate when varying IoU threshold for Top k object proposals on test set images. SalProp is the best technique at lower number of proposals achieving over 25% and 19% recall with only 1 window at IoU = 0.5 and 0.6 respectively. At IoU = 0.7, SalProp outperforms Rahtu [42] by 3.46%, Selective Search (SS) [11] by 5.16%, Objectness (OBJ) [2] by 7.32%, Randomized Prim's (RP) [43] by 8.71%, GOP [46] by 22.36%, Rigor [45] by 23.46%, Rantalankila (Ranta) [44] by 30.05% and Perceptual Edge (PE) [4] by 30.35% at top-10 proposals demonstrating that it consistently ranks higher the object proposals that are closer to the ground truth when lower number of proposals are considered.

MCG [12] and EdgeBoxes70 (EB) [3] techniques outperform SalProp technique. MCG has a hierarchical segmenter that makes effective

Table 1
Performance metrics used in seam carving evaluation.

Performance metrics	Relevance
SIFT Flow[31]	SIFT Flow helps in computing smoothened flow fields using SIFT descriptors of the two images.
Earth Mover's Distance (EMD)[32]	The Earth Mover's Distance (EMD) [29] is defined as the minimal cost incurred when one histogram P is transformed to another histogram Q, given a ground distance metric between the basic features that are combined to form the histogram.
Aspect Ratio Similarity (ARS)[33]	The block pair of the regular partitioned block $N \times N$ in the original image and the corresponding block in retargeted image is utilized to calculate the local aspect ratio similarity scores. When the ARS score is close to 1, the block content in original image is generally kept in high quality in retargeted image, while when ARS score is close or equal to zero, it indicates that the retargeted block is suffering from serious information loss and distortion or even removed totally.
Salient Global Structure Distortion Measurement (SGSD) [34]	It evaluates the measurement of total structure distortion.
Forward Saliency Information Loss (FSIL) [34]	It denotes the amount of salient information retained after performing a retargeting operation. It ranges between [0–1].
Backward Saliency Information Loss (BSIL) [34]	To compute the backward salient information loss content, this index is used which quantifies the amount of salient information that can be recovered from a retargeted image. Using the retargeted image, an image equivalent to original image is recovered using SIFT-flow which is used to compute the BSIL metric ranging between [0–1].

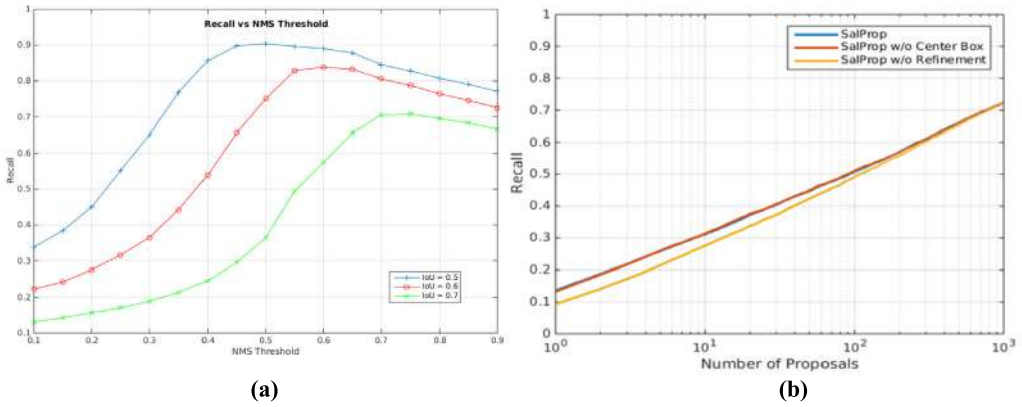

Fig. 5. (a) NMS cut-off threshold for highest recall value at varying IoU values and (b) SalProp algorithm variant (SalProp7) at IoU = 0.7 showing the detection accuracy with and without removal of edges in the center-box and bounding box location refinement on validation set images.

use of multiscale information and a grouping engine to group the multiscale regions into object proposals. The results demonstrate that the proposed algorithm performs better on varying IoU thresholds for less number of candidate proposals while maintaining high recall at higher proposals as compared to rest other techniques. The important note to make here is that except Objectness the compared approaches do not take into account the saliency aspect of an object which is a key property in characterizing an object. Objectness relies on explicit multi-scale saliency map to localize the key objects on the other hand SalProp technique determines the saliency value of the edge segments to rank the key objects according to their relative order of salience. Although CRF training is conducted on a saliency dataset, the generalization ability of the trained model is shown on a generic object dataset PASCAL VOC 2007 and 2012 datasets. Thus, the technique is not specific to saliency datasets. Our method consistently outperforms objectness by 2%, 6% and 30% at IoU thresholds 0.5, 0.6 and 0.7 respectively. In contrast to EdgeBoxes approach, which performs edge grouping based on similar orientation, we adopt a computationally efficient CRF learning based setting.

Retrieval Time. SalProp provides significant performance gain over other methods (PE, OBJ, Rahtu, RP, Ranta, SS, Rigor, GOP) while lagging performance to EdgeBoxes and MCG. It achieves a computational speedup of 5x over MCG (Table 2). SalProp does not require any training on region-level salient object ground truth data as it is based on edge saliency map construction. However, the point to note here is that EdgeBoxes and MCG does not incorporate the saliency ranking aspect of object proposals.

5.2.1. Seam carving quantitative results

In Table 3 we compare the traditional seam carving methods and image retargeting methods using aforementioned IRQA metrics. For a fair evaluation we also report the performance of other content aware seam carving methods (SC [27]) and saliency map guided seam carving (SC-GBVS [28,36], SC-DRFI [28,37], SC-DEEPSAL [28,38], SC-PDnet [28,40], SC-mDRFI [28,41]). We have also shown the evaluation on the proposed seam carving (CASC) with other compared object proposal generation techniques (CASC-EdgeBoxes and CASC-MCG). In saliency map guided seam carving, the saliency map serves as the importance map to remove the seams from regions with least importance (saliency values) thus not hindering the overall structural similarity of the salient objects in the final retargeted image. However, the saliency map does not cover the objects which might not be as salient. Since, the proposed method (CASC) is not relying on any explicit saliency map computation rather it works on key object ranking based on object edge density it works on the principle of generic object detection. We report the average values obtained on these metrics on the MSRA1000 saliency dataset [30]. It can be observed that the proposed method (CASC) achieves almost equivalent average $SIFT_{Flow}$ distance as compared to linear scaling (SCL) and SNS while almost an order lesser than traditional seam carving methods (SC and SCMIT), saliency based seam carving (SC-GBVS, SC-DRFI, SC-DEEPSAL, SC-PDnet, SC-mDRFI) and cropping. Similar analogies can be inferred for the EMD metric as well showing that the content aware seam carving results have much more similarity score to the original image. In ARS metric, for 25% and 50% reduction we achieve far higher ARS index than traditional seam

Table 2
Comparison of top 1000 proposals on Pascal VOC 2007 and 2012 datasets with state-of-the-art techniques on AUC% (higher the better), number of proposals (N) at 75% recall (lower the better) and recall% (higher the better). '–' indicates that the particular recall rate is not reached. Run-time analysis of various methods is also shown. The value in bold indicates the best results.

PASCAL VOC 2007										
Method	IoU = 0.5			IoU = 0.6			IoU = 0.7			Time
	AUC	N @75%	Recall	AUC	N @75%	Recall	AUC	N @75%	Recall	
EB70 [3]	65.82	86	93.45	60.52	141	**90.73**	53.03	**294**	**84.15**	0.25
PE [4]	1.8	–	10.4	0.08	–	4.7	0.02	–	1.2	7.2
MCG [12]	**71**	**37**	**94.6**	**62.8**	**95**	90.2	**62.5**	366	83	34
OBJ [2]	62	145	89	52	504	78	30	–	41	3
Rahtu [42]	57	278	84	50	551	79	43.5	–	73.5	3
RP [43]	59.3	129	89	50	315	83	40.7	1000	75	1
Ranta [44]	25.14	511	86.38	21.63	718	79.77	17.76	–	70.75	10
SS [11]	62.3	105	93	54	207	88	45.3	544	80	10
Rigor [45]	40.39	–	67.43	32.05	–	54.5	23.44	–	40.73	6.84
GOP [46]	47.8	155	93	41	272	87	33.4	705	76	0.9
SalProp	67.5	74	91	58.1	244	84	44	–	71.3	7.11
PASCAL VOC 2012										
Method	IoU = 0.5			IoU = 0.6			IoU = 0.7			Time
	AUC	N @75%	Recall	AUC	N @75%	Recall	AUC	N @75%	Recall	
EB70 [3]	92.29	56	**98.22**	**88.68**	94	**96.68**	**80.62**	**195**	**90.46**	0.25
PE [4]	6.96	–	15.17	4.01	–	9.39	1.33	–	3.08	7.2
MCG [12]	**92.93**	**30**	97.3	86.75	**73**	93.17	76.53	253	84.78	34
OBJ [2]	90.84	67	96.74	79.31	177	86.23	40.26	–	42.79	3
Rahtu [42]	80.08	227	90.47	72.92	408	85.04	64.16	789	78.06	3
RP [43]	81.27	206	92.18	69.32	528	83.16	54.39	–	69.31	1
Ranta [44]	30.05	–	63.64	16.47	718	40.74	7.4	–	20.83	10
SS [11]	88.62	98	96.41	79.55	227	90.04	66.34	709	78.23	10
Rigor [45]	19.96	–	33.78	8.88	–	15.62	3.07	–	5.65	6.84
GOP [46]	75.91	298	95.13	60.62	629	83.69	40.30	–	61.51	0.9
SalProp	92.41	56	98.18	84.54	143	93.74	68.58	588	80.78	7.11

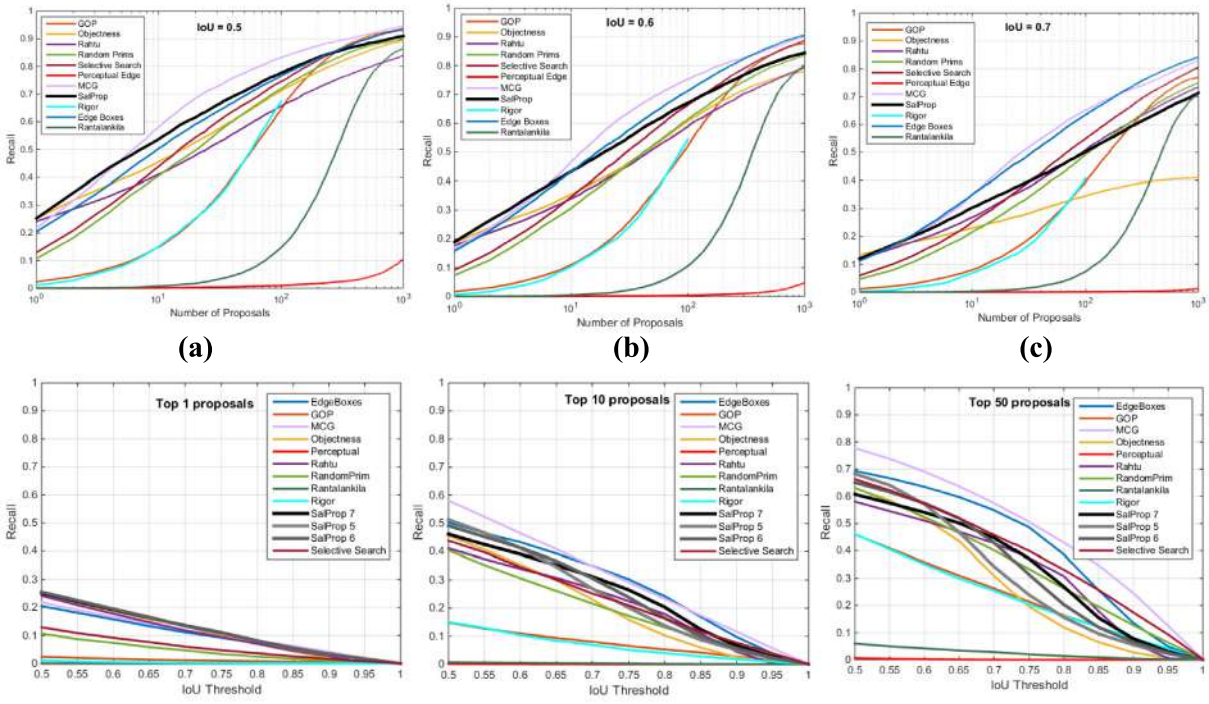

Fig. 6. (a)–(c) The detection rate vs. the number of bounding box proposals for varying IoU = 0.5, 0.6 and 0.7 and (d)–(f) The detection rate vs. intersection over union for Top k object proposals on test set images. The variations of SalProp framework are tested using IoU threshold = 0.5, 0.6 and 0.7 indicated by SalProp 5, SalProp 6 and SalProp 7 respectively on Pascal VOC 2007 4952 test images.

carving techniques SC, SCMIT, SC-GBVS, SC-DRFI, SC-DEEPSAL, SCC-PDnet, SC-mDRFI whereas slightly lagging behind linear scaling (SCL) and SNS. This validates that the aspect ratio is maintained with the proposed seam carving method to much higher extent which is not the case with generic seam carving techniques. The saliency information loss metrics $FSIL$ and $BSIL$ are quite high than other compared seam carving methods in the proposed method (both for 25% and 50% reduction) indicating that content information loss is minimalistic.

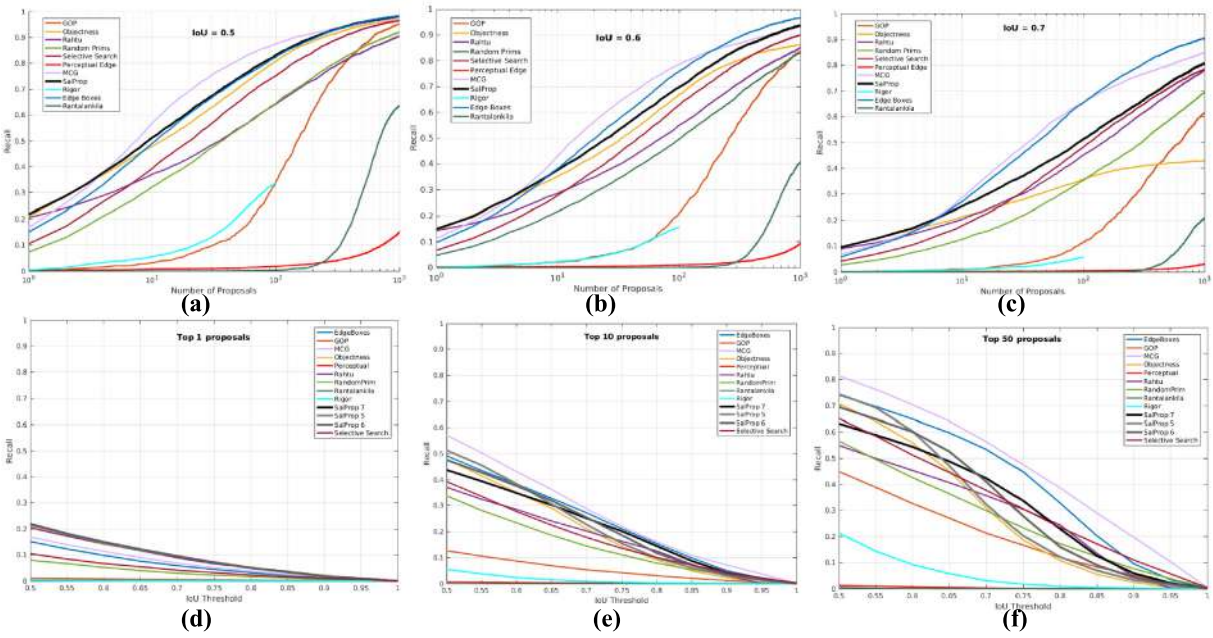

Fig. 7. (a)–(c) The detection rate vs. the number of bounding box proposals for varying IoU = 0.5, 0.6 and 0.7 and (d)–(f) The detection rate vs. intersection over union for Top k object proposals on test set images. The variations of SalProp framework are tested using IoU threshold = 0.5, 0.6 and 0.7 indicated by SalProp 5, SalProp 6 and SalProp 7 respectively on Pascal VOC 2012 5138 test images.

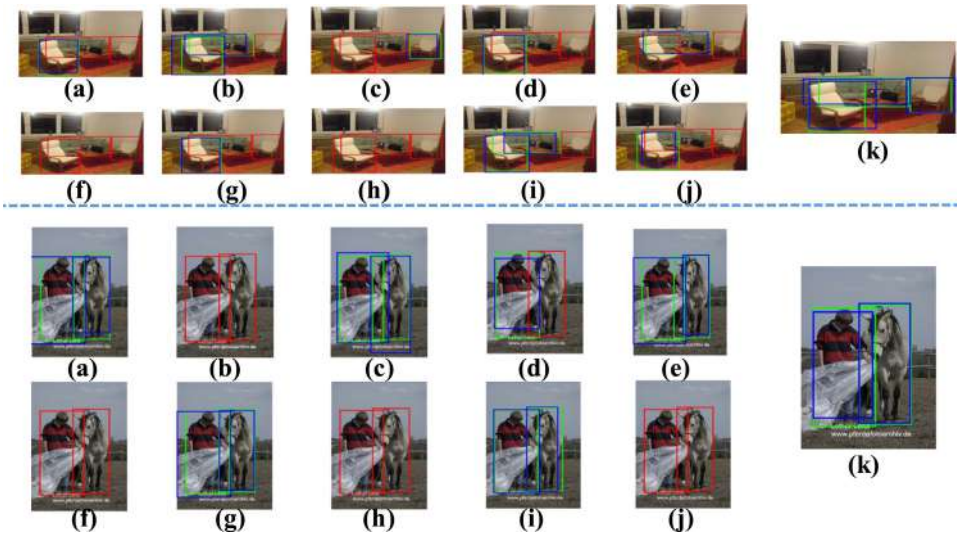

Fig. 8. Qualitative examples of our object proposals with other competing techniques. (a) GOP (b) OBJ (c) Rahtu (d) RP (e) SS (f) PE (g) MCG (h) Rigor (i) EB70 (j) Ranta (k) SalProp. Blue bounding boxes are the closest produced object proposals to each ground truth bounding box shown in green. Missed objects are shown with bounding boxes indicated in red meaning that the object was not found. IoU threshold = 0.7 was used to determine correctness for all examples.

We are able to achieve an average gain of 14.22% at 25% reduction and 7.8% at 50% reduction for $FSIL$ metric over the compared seam carving methods although we lag behind CR (cropping) by a margin of 9.8% at 25% reduction and 18.05% at 50% reduction and behind SC-GBVS by an average gap of 4.31%. For $BSIL$ metric, we achieve the best results in both cases. We also achieve lesser distortion rates ($SGSD$ metric) than the reported seam carving methods. In comparison to the performance of CASC-EdgeBoxes and CASC-MCG, the proposed approach obtains significant performance gains over various IRQA metrics. The reason for this can be attributed to the salience ranking of proposal set in CASC-SalProp.

5.3. Qualitative evaluation

Fig. 8 shows qualitative examples of SalProp proposals and other object proposal techniques. The results are computed for IoU = 0.7. It can be observed that SalProp produces tight bounding boxes (e.g. sofa sets and horses) and is able to detect occluding and difficult objects with high accuracy.

Effect of varying edge parameter settings. Fig. 9(a)–(c) shows the experimental analysis with varying edge length and magnitude. In case of low magnitude threshold there are a lot of spurious edges and in high threshold it results in loss of object edges. The same inference is observed in case of variation with edge lengths. Fig. 9(c) shows the

Table 3
Performance evaluation using IRQA metrics: (a) SIFT-Flow (Lower the better) (b) Earth Mover's Distance (EMD) (Lower the better) (c) Aspect Ratio Similarity (ARS) (Higher the better) (d) Salient Global Structure Distortion Measurement (SGSD)(Lower the better) (e) Forward Saliency Information Loss (FSIL) (Higher the better) (f) Backward Saliency Information Loss (BSIL) (Higher the better). Redn. = Reduction/Carving%. The value in bold indicates the best results.

Method	Redn.	SIFT_Flow	EMD	ARS	SGSD	FSIL	BSIL
SC[27]	**25%**	1.3711e+03	801.4917	0.9025	0.9541	0.7073	0.9777
	50%	2.4822e+03	1.1554e+03	0.7934	0.928	0.4226	0.9249
SCMIT[28]	**25%**	1.3357e+03	799.8052	0.8953	0.9523	0.6996	0.981
	50%	2.4294e+03	1.1529e+03	0.7865	0.923	0.4161	0.9345
SCL	**25%**	**1.1717e+03**	**777.4367**	**0.9637**	0.9394	0.5693	0.9871
	50%	**2.0803e+03**	1.1392e+03	0.877	0.9018	0.3208	0.9645
CR	**25%**	1.2177e+03	785.5555	0.9344	0.9772	**0.9055**	0.979
	50%	2.2911e+03	**1.1334e+03**	0.8592	0.9767	**0.563**	0.8567
SNS[39]	**25%**	1.1771e+03	777.8810	0.9631	0.9392	0.569	0.9869
	50%	2.0816e+03	1.1393e+03	**0.8779**	**0.9012**	0.321	0.9643
SC-GBVS[28,36]	**25%**	1.3122e+03	788.9545	0.9245	0.9873	0.8350	0.9746
	50%	2.3425e+03	1.1468e+03	0.8346	0.9434	0.5117	0.9158
SC-DRFI[28,37]	**25%**	1.2914e+03	803.3900	0.9137	0.9776	0.7572	0.9638
	50%	2.1415e+03	1.1538e+03	0.8393	0.9427	0.4701	0.8930
SC-DEEPSAL[28,38]	**25%**	1.2742e+03	804.6370	0.8963	0.9691	0.7229	0.9613
	50%	2.1022e+03	1.1526e+03	0.8300	0.9448	0.4642	0.8930
SC-PDnet[28,40]	**25%**	1.2664e+03	802.1200	0.9361	0.9519	0.7329	0.9712
	50%	2.0870e+03	1.1432e+03	0.8465	0.9343	0.4766	0.9030
SC-mDRFI[28,41]	**25%**	1.2857e+03	803.7000	0.9200	0.9772	0.7476	0.9642
	50%	2.1302e+03	1.1530e+03	0.8411	0.9420	0.4684	0.9019
CASC	**25%**	1.1765e+03	778.8652	0.9467	**0.9329**	0.824	**0.9872**
	50%	2.0813e+03	1.1395e+03	0.8737	0.9013	0.4769	**0.9675**
CASC-EdgeBoxes[3]	**25%**	1.3789e+03	804.3811	0.9022	0.9560	0.7229	0.9799
	50%	2.4942e+03	1.1587e+03	0.7922	0.9328	0.4477	0.9208
CASC-MCG[12]	**25%**	1.3531e+03	801.0967	0.9140	0.9706	0.7769	0.9815
	50%	2.4330e+03	1.1503e+03	0.8424	0.9368	0.4539	0.9388

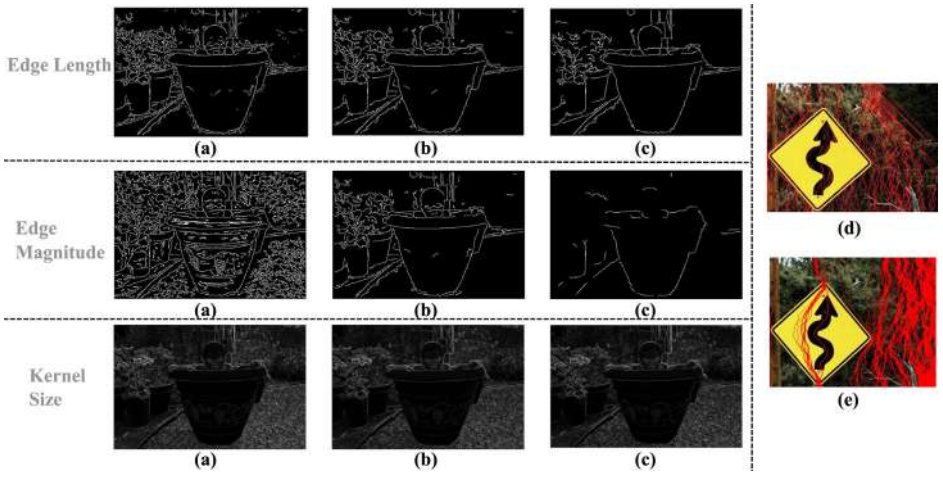

Fig. 9. Validation Experiments by varying edge parameters: edge length (Top Row), edge magnitude (Middle Row) and kernel size (Bottom Row) for computation of color gradients along the edge segment. (a)–(c) indicates Low, Medium and High Threshold ranges. Kernel Sizes are 3, 5, 7 respectively. Seam visualization: (d) SalProp object proposals guided seam carving (e) SCMIT.

filtering results with varying kernel sizes. As the kernel size increases, there is loss in the subtle texture changes.

5.3.1. Seam carving qualitative results

As can be seen in Fig. 10, the proposed seam carving strategy is able to preserve the geometric structure of the salient objects whereas traditional seam carving approaches (SC, SCMIT) completely lose the overall structure of the object. In cropping (CR) results the object is cut out (e.g. red box and clock tower). In Fig. 11, proposed seam carving technique results with multi-aspect ratio targets is shown.

Impact of content aware seam carving. In Fig. 9(d)–(e), we compare the seam carving guided by SalProp object proposals to other seam carving approaches. It can be clearly seen that lesser number of seams pass through the object in our method than in the existing technique SCMIT due to the notion of object protection we consider preserving the overall aesthetics of the image.

6. Conclusion

We have proposed a novel salient object proposal generation technique which utilizes salient object edge density. We obtained prioritized tight proposals and thus perform effective content aware seam carving. We have exhaustively performed empirical analysis with several baseline methods to show the effectiveness of the proposed architecture at varying IoU thresholds. We have also demonstrated the object discriminability of the proposed object proposal strategy in automated

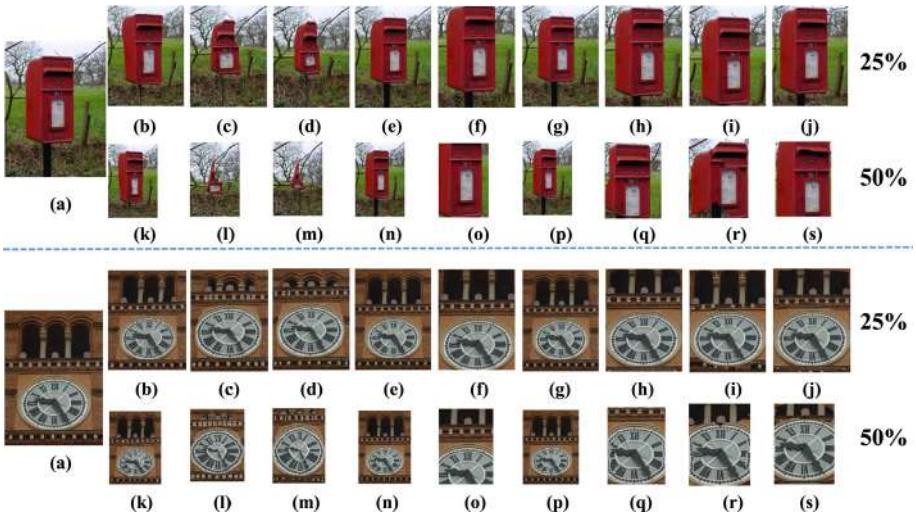

Fig. 10. Qualitative examples of seam carving using SalProp based object hypotheses at 25% and 50% carving results. (a) Original image (b) Seam Carving using SalProp object hypotheses (c) SC (d) SCMIT (e) SNS (f) CR (g) SCL (h) SC-GBVS (i) SC-DRFI (j) SC-DEEPSAL at 25% reduction. (k) Seam Carving using SalProp object hypotheses (l) SC (m) SCMIT (n) SNS (o) CR (p) SCL (q) SC-GBVS (r) SC-DRFI (s) SC-DEEPSAL at 50% reduction.

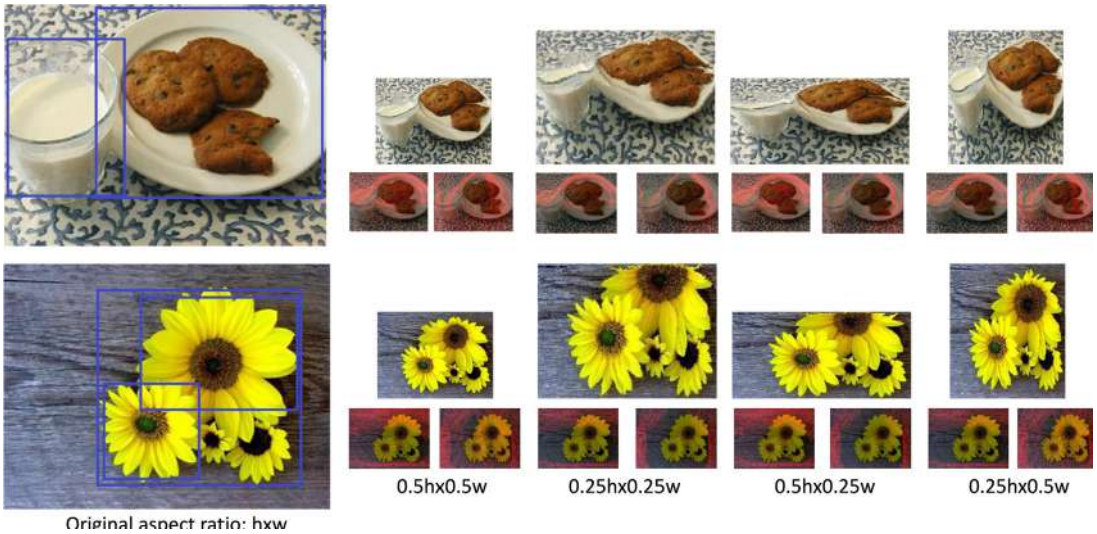

Fig. 11. CASC technique's seam carving results on multiple objects at multi-aspect ratio targets. Seam visualization shows the horizontal seams (left) and vertical seams (right) removed overlaid on original image. Best boxes using SalProp are also shown.

content aware seam carving to protect the key objects. In future, we intend to incorporate depth cue and extend the work for RGBD images and videos.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.
[2] B. Alexe, T. Deselaers, V. Ferrari, Measuring the objectness of image windows, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2189–2202.
[3] C.L. Zitnick, P. Dollár, Edge boxes: Locating object proposals from edges, in: European Conference on Computer Vision, Springer, 2014, pp. 391–405.
[4] Y. Qi, Y.-Z. Song, T. Xiang, H. Zhang, T. Hospedales, Y. Li, J. Guo, Making better use of edges via perceptual grouping, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1856–1865.
[5] Y. Yan, J. Ren, G. Sun, H. Zhao, J. Han, X. Li, S. Marshall, J. Zhan, Unsupervised image saliency detection with gestalt-laws guided optimization and visual attention based refinement, Pattern Recognit. 79 (2018) 65–78.
[6] P. Mukherjee, B. Lall, S. Tandon, Salprop: Salient object proposals via aggregated edge cues, in: Image Processing (ICIP), 2017 IEEE International Conference on, IEEE, 2017, pp. 2423–2429.
[7] C. Xia, F. Qi, G. Shi, P. Wang, Nonlocal center–surround reconstruction-based bottom-up saliency estimation, Pattern Recognit. 48 (4) (2015) 1337–1348.
[8] X. Zhang, C. Xu, M. Li, R.K. Teng, Study of visual saliency detection via nonlocal anisotropic diffusion equation, Pattern Recognit. 48 (4) (2015) 1315–1327.
[9] L. Huo, L. Jiao, S. Wang, S. Yang, Object-level saliency detection with color attributes, Pattern Recognit. 49 (2016) 162–173.
[10] M.-M. Cheng, N.J. Mitra, X. Huang, P.H. Torr, S.-M. Hu, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2015) 569–582.
[11] J.R. Uijlings, K.E. van de Sande, T. Gevers, A.W. Smeulders, Selective search for object recognition, Int. J. Comput. Vis. 104 (2) (2013) 154–171.
[12] J. Pont-Tuset, P. Arbelaez, J.T. Barron, F. Marques, J. Malik, Multiscale combinatorial grouping for image segmentation and object proposal generation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (1) (2017) 128–140.
[13] F. Shafieyan, N. Karimi, B. Mirmahboub, S. Samavi, S. Shirani, Image retargeting using depth assisted saliency map, Signal Process., Image Commun. 50 (2017) 34–43.
[14] Y. Guo, Y. Hao, M. Yu, Image retargeting quality assessment based on content deformation measurement, Signal Process., Image Commun. 67 (2018) 171–181.

[15] Y. Lyu, S.-P. Lu, Q. Bolsee, A. Munteanu, Consistent video projection on curved displays, Signal Process., Image Commun. (2019).
[16] J. Hosang, R. Benenson, P. Dollár, B. Schiele, What makes for effective detection proposals?, IEEE Trans. Pattern Anal. Mach. Intell. 38 (4) (2016) 814–830.
[17] W. Li, H. Li, B. Luo, H. Shi, Q. Wu, K.N. Ngan, Improving object proposals with top-down cues, Signal Process., Image Commun. 56 (2017) 20–27.
[18] X. Chen, H. Ma, X. Wang, Z. Zhao, Improving object proposals with multi-thresholding straddling expansion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2587–2595.
[19] Z. Fang, Z. Cao, Y. Xiao, L. Zhu, J. Yuan, Adobe boxes: Locating object proposals using object adobes, IEEE Trans. Image Process. 25 (9) (2016) 4116–4128.
[20] W. Kuo, B. Hariharan, J. Malik, Deepbox: Learning objectness with convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2479–2487.
[21] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. (6) (1986) 679–698.
[22] J.J. Lim, C.L. Zitnick, P. Dollár, Sketch tokens: A learned mid-level representation for contour and object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3158–3165.
[23] C. Aytekin, A. Iosifidis, M. Gabbouj, Probabilistic saliency estimation, Pattern Recognit. 74 (2018) 359–372.
[24] S. Hallman, C.C. Fowlkes, Oriented edge forests for boundary detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1732–1740.
[25] P. Dollár, C.L. Zitnick, Fast edge detection using structured forests, IEEE Trans. Pattern Anal. Mach. Intell. 37 (8) (2015) 1558–1570.
[26] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Trans. Image Process. 19 (6) (2010) 1635–1650.
[27] S. Avidan, A. Shamir, Seam carving for content-aware image resizing, in: ACM Transactions on Graphics (TOG), Vol. 26, ACM, 2007, p. 10.
[28] M. Rubinstein, A. Shamir, S. Avidan, Improved seam carving for video retargeting, in: ACM Transactions on Graphics (TOG), Vol. 27, ACM, 2008, p. 16.
[29] A.C. Müller, S. Behnke, Pystruct: learning structured prediction in python., J. Mach. Learn. Res. 15 (1) (2014) 2055–2060.
[30] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: Computer Vision and Pattern Recognition, 2009. Cvpr 2009. Ieee Conference on, IEEE, 2009, pp. 1597–1604.
[31] C. Liu, J. Yuen, A. Torralba, J. Sivic, W.T. Freeman, Sift flow: Dense correspondence across different scenes, in: European Conference on Computer Vision, Springer, 2008, pp. 28–42.
[32] O. Pele, M. Werman, Fast and robust earth mover's distances, in: Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 460–467.
[33] Y. Zhang, Y. Fang, W. Lin, X. Zhang, L. Li, Backward registration-based aspect ratio similarity for image retargeting quality assessment, IEEE Trans. Image Process. 25 (9) (2016) 4286–4297.
[34] Z. Chen, J. Lin, N. Liao, C.W. Chen, Full reference quality assessment for image retargeting based on natural scene statistics modeling and bi-directional saliency similarity, IEEE Trans. Image Process. 26 (11) (2017) 5138–5148.
[35] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[36] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Advances in Neural Information Processing Systems, 2007, pp. 545–552.
[37] J. Wang, H. Jiang, Z. Yuan, M.-M. Cheng, X. Hu, N. Zheng, Salient object detection: A discriminative regional feature integration approach, Int. J. Comput. Vis. (ISSN: 1573-1405) 123 (2) (2017) 251–268, http://dx.doi.org/10.1007/s11263-016-0977-3.
[38] X. Li, L. Zhao, L. Wei, M.-H. Yang, F. Wu, Y. Zhuang, H. Ling, J. Wang, Deepsaliency: multi-task deep neural network model for salient object detection, IEEE Trans. Image Process. 25 (8) (2016) 3919–3930.
[39] Y.-S. Wang, C.-L. Tai, O. Sorkine, T.-Y. Lee, Optimized scale-and-stretch for image resizing, in: ACM Transactions on Graphics (TOG), Vol. 27, ACM, 2008, p. 118.
[40] C. Zhu, X. Cai, K. Huang, T.H. Li, G. Li, Pdnet: Prior-model guided depth-enhanced network for salient object detection, in: 2019 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2019, pp. 199–204.
[41] M. Jahanifar, N.Z. Tajeddin, B.M. Asl, A. Gooya, Supervised saliency map driven segmentation of lesions in dermoscopic images, IEEE J. Biomed. Health Inf. 23 (2) (2018) 509–518.
[42] E. Rahtu, J. Kannala, M. Blaschko, Learning a category independent object detection cascade, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 1052–1059.
[43] S. Manen, M. Guillaumin, L. Van Gool, Prime object proposals with randomized Prim's algorithm, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2536–2543.
[44] P. Rantalankila, J. Kannala, E. Rahtu, Generating object segmentation proposals using global and local search, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2417–2424.
[45] A. Humayun, F. Li, J.M. Rehg, Rigor: Reusing inference in graph cuts for generating object regions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 336–343.
[46] P. Krähenbühl, V. Koltun, Geodesic object proposals, in: European Conference on Computer Vision, Springer, 2014, pp. 725–739.