PoS PROCEEDINGS OF SCIENCE

Foreground Object Segmentation with Objectness Measure

Junguang Zhu[1]
National Digital Switching System Engineering & Technological Research Center
Zhengzhou, 450002,China
E-mail: adamzhu1990@163.com

lixin Ji[2]
National Digital Switching System Engineering & Technological Research Center
Zhengzhou, 450002, China
E-mail: jlxndsc@163.com

chengqiu xiao[3]
Troop 75753, People's Liberation Army
Guangzhou, 510600, China
E-mail: xiaochengqiu123@163.com

This paper proposes a novel model to address the problem of image segmentation with objectness measure. Recently, many objectness measures are proposed, which aims to generate candidate windows to localize the possible objects in the image. Consider to combine this useful object location piror into a foreground segment model. Specifically, a Conditional Random Filed model is constructed on superpixels graph, and it efficiently incorporates objectness measure, color distribution and appearance similarity. Expermental results on a extended GrabCut dataset demonstrate that the proposed model can yield a foreground object segmentation of better quality.

ISCC2015
18-19, December, 2015
Guangzhou, China

[1]Speaker
[2]Correspongding Author
[3]This study is supported by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 61521003)

PoS(ISCC2015)039

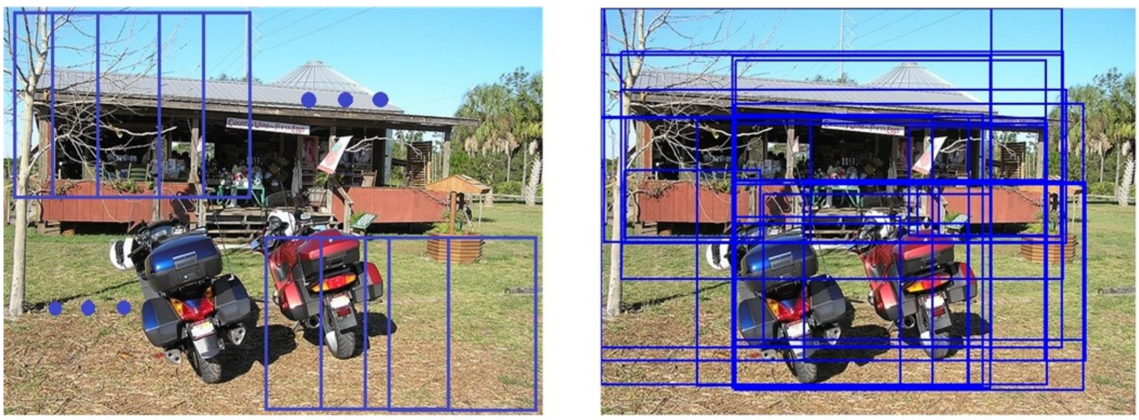
1. Introduction

Foreground object segmentation is a technique for extracting a foreground region in an image from its background. The goal is a general purpose that divides an image into two segments: "foreground " and " background". However, the fully automatic segmentation seems not to be perfect. For this reason, usually it is needed to impose some seeds as constraints, which indicate the part of region belonging to foreground/background. According to the similarities between the seeds and other unlabeled regions, they can easily perform better segmentation. Usually, the object segmentation can also be used in many tasks, including object recognition and object categorization. As segmentation can be regarded as a labeling problem, it ususally relys on a specific CRF/MRF to label each pixle as foreground or background. The CRF/MRF is a kind of probabilistic graphical model, which can infer the foreground likelihood under given observations.

On the one hand, among the observations, mostly attention is paid to describe foreground object appearance without exploiting object location information; on the other hand, the recently proposed objectness measure[1-3], which generates a set of scored windows to measure the object existent extent, can exactly provide the location prior. In order to exploit this useful location prior, we innovatively combine the objectness measure into a CRF model for a better segmentation.

2. Objectness Measure

The objectness measure is alternative framework for sliding windows, and it is widely used in current object detection methods[4]. While sliding windows need to exhaustively search around $10^5 \sim 10^6$ windows per image to locate foreground object, objectness measure only generates a few number of windows, as shown in Fig. 1. The objectness measure is based on a reasonable assumption that all categories of objects share some common visual properties, which can distinguish them from the background. The windows generated by objectness measure often lie around the foreground object with confidence scores. The confidence scores can represent how likely there is an object in the windows. So, the density of spatial distribution for all objectness windows and the sum of confidence scores can describe the positions of the objects.



(a). silding windows (b). objectness measure

Figure 1: (a) Sliding Windows Search V.S. (b) Objectness Measure

The conception of "Objectness" was first clarified by Alexe et al. as a generic measure, evaluating how likely it is for a window to contain an object of any class [1]. They, at the earliest, realized the measure through combining multiple cues into a Bayesian framework to filter the sampled windows. The initial windows are sampled from dense regular grids, according to the distribution of the saliency scores for all windows in the grids. Base on the same thought, Rahtu et al. proposed the variation[2]. They modified the objectness features, based on a max-margin structured frameworkto learn feature combinations[5]. Recently, *EdgeBoxes*observes that the objects are always surrounded by complete boundary[3], and the boundary density in windows is a powerful objectness feature to locate the objects. So, based on the characters of the continuity over boundary orientation and the integrality of close boundary, they extract boundary information from the edge map.

3. Algorithm Pipline

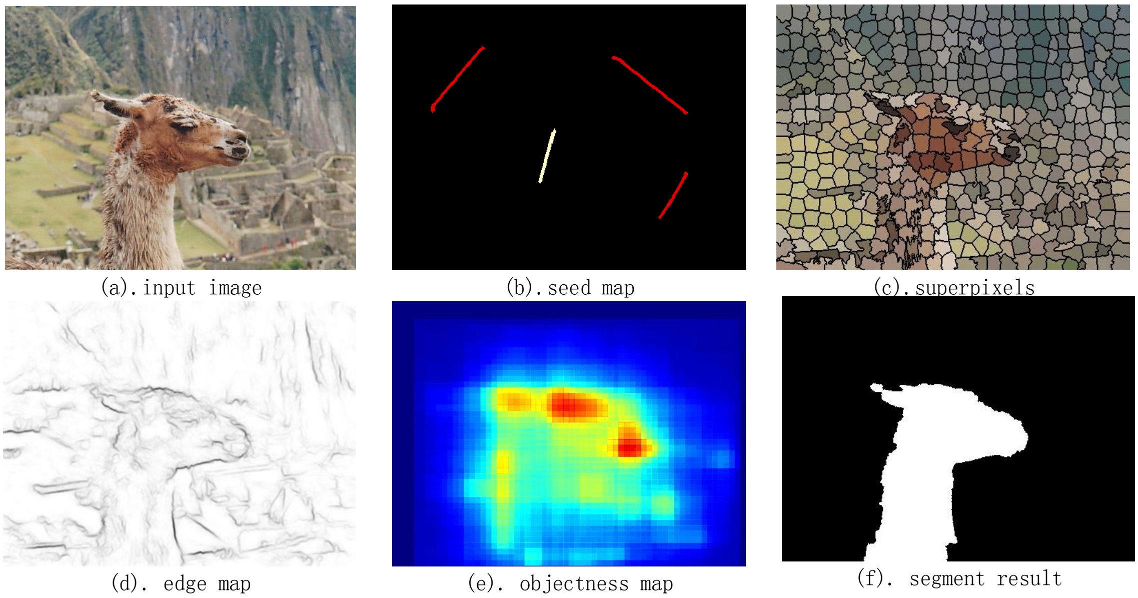


Figure 2: Algorithm Pipline

The algorithm pipline is showed in Fig. 2. Given the input image (a) and correspongding seed map (b), first, we use the *SLIC* methodto oversegment the input image into several superpixels (c)[6], instead of directly using pixels for efficiency reason. Each superpixel has uniform color distribution and can keep good object boundaries; then, we extract the edge map (d), and use *EdgeBoxes* to measure the objectness over input image to get an objectness map (e). According to the seed map, estimate the foreground/background color distribution. Adjacent superpixels' similarities are evaluated on the edge map; finally, combining the objectness measure, color distribution, and appearance similarity into a CRF model to segment foreground object (f).

4. Foreground Segment Model Based on CRF

Let $X=\{x_i\}_{i=1}^{n}$ be the superpixels collection of an image, with corresponding binary label $Y=\{y_i\}_{i=1}^{n}$. Each superpixel x_i is either labeled as foreground or background, according to $y_i=0$ or 1. The conditional probability of foreground labeling Y gave an image X as

$$P(Y|X,W)=\frac{1}{Z}e^{-E(X,Y;W)} \tag{4.1}$$

Where W are weights, Z is the partition function. $E(Y,X;W)$ is the energy function, which is defined as a sum of all penalization under specific labels.

$$P(Y|X,W)\propto -E(X,Y;W) \tag{4.2}$$

The conditional probability $P(Y|X, W)$ is negative correlation with energy function. In the conditional probability, to infer the maximum a posteriori (MAP) label *Y*, is equivalent to minimize the energy function. For each image, a graph $G=(\nu, \varepsilon)$ is created, which has a set of nodes, ν (individual superpixel) and edges, ε (pairs of adjacent superpixels). The energy function $E(Y, X; W)$ is the linear combination of the following three potential functions:

$$E(Y, X; W)=\sum_{i\in\nu}\psi_i(y_i; x_i, w_{objness})+\sum_{i\in\nu}\gamma_i(y_i; x_i, w_{color})+\sum_{(i,j)\in\varepsilon}\phi_{i,j}(y_i, y_j; x_i, x_j, w_{similar}) \tag{4.3}$$

In the energy function, the first trem $\psi_i(y_i; x_i, w_{objness})$ is a objectness unary potential, which provides the objectness measure of superpixels. The second trem $\gamma_i(y_i; x_i, w_{color})$ is a color unary potential, which according to each node's color distribution computes the foreground/background likelihood. The third trem $\phi_{i,j}(l_i, l_j; x_i, x_j, w_{similar})$ is a similarty pairwise potential, which according to appearance similarty creates the mutual relation between adjacent nodes.

4.1 Objectness Unary Potential

The objectness unary potiential evaluates the likelihood that a superpixel belongs to an object. It is based on the objectness measure *EdgeBoxes* to generate a set of windows with objectness scores. Accumulating all the scores in each superpixel, and normalizing them with maximum score in image can get an objectness map. The objectness unary potential is defined on the objectness map as follow:

$$\psi_i(y_i; x_i, w_{objness})=w_{objness}\cdot f_{obj}(y_i, x_i) \tag{4.4}$$

where

$$f_{obj}(y_i, x_i)=\begin{cases}\dfrac{\exp(-\alpha\cdot s(x_i))}{1+\exp(-\alpha\cdot s(x_i))}, & y_i=1\\ \dfrac{1}{1+\exp(-\alpha\cdot s(x_i))}, & y_i=0\end{cases} \tag{4.5}$$

$f_{obj}(y_i, x_i)$ is a smooth mapping function based on the exponent form, which builds the relation between objectness score $s(x_i)$ and different label y_i. When $y_i=1$, it penalizes the superpixel with low objectness score, which is labeled as foreground. Otherwise, when $y_i=0$, it penalizes the superpixel with high objectness score, which is labeled as background. According to the distribution of objectness scores, the coefficient α can adjust the smooth of maping function. Here we set $\alpha=2.5$ as an optimal coefficient.

4.2 Color Unary Potential

The color unary potential is based on the contrast of color distribution between the seed superpixels and unlabeled superpixels. In order to obtain the color distribution within each superpixel, compute the color histogram $\{h_L, h_a, h_b\}$ in Lab color space. There, each color channel is quantized as 45 bins. First, according to the L1 distance of color histograms between the foreground/background seed superpixels and unlabeled superpixels x_i, the foreground/background probability distribution is defined as:

$$P_f(x_i)=\exp\left(-\lambda\cdot\sum_{c\in\{L,a,b\}}\|h_c(x_i)-h_c(x_f)\|\right) \tag{4.6}$$

$$P_b(x_i)=\exp\left(-\lambda\cdot\sum_{c\in\{L,a,b\}}\|h_c(x_i)-h_c(x_b)\|\right) \tag{4.7}$$

Where λ is a scaling factor; then the color contrast function $f_{color}(y_i, x_i)$ is based on the probability distribution

$$f_{color}(y_i, x_i)=\ln\ P_f(x_i)-\ln\ P_b(x_i) \tag{4.8}$$

And the color unary potential can be computed as:

$$\gamma_i(y_i; x_i, w_{color})=w_{color}\cdot F_{color}(y_i, x_i) \tag{4.9}$$

Where

$$F_{color}(y_i, x_i)=\begin{cases} 0 & if \quad y_i=1, \quad x_i\notin x_b \\ \infty & if \quad y_i=1, \quad x_i\in x_b \\ \infty & if \quad y_i=0, \quad x_i\in x_f \\ f_{color}(y_i, x_i) & if \quad y_i=0, \quad x_i\notin x_f \end{cases} \tag{4.10}$$

When the superpixel x_i has the similar color distribution with foreground seed superpixels x_f, $P_b(x_i)$ has a samll value, while $P_f(x_i)$ is larger, which causes the color contrast function $f_{color}(y_i, x_i)$ having a larger value. So, if this superpixel x_i is labeled as background, for $y_i=0$, the penalization is larger. As the foreground seed superpixels are labeled as background, the penalization is infinity.

4.3 Similarity Pairwise Potential

The similarity pairwise potential $\phi_{i,j}$ models the intersection between the two labels y_i and y_j of two neighboring superpixels based on the appearance similarity, and this pairwise term has a spatial smooth effection on the labels of neighboring superpixels. To measure the similarity between the siperpixels, we compute a kind of normalized histogram descriptor h. The histogram descriptor is composed of color and texture histograms h_c and h_t. Using the histogram intersection distance, we can extract the similarity between a pair of adjacent superpixels x_i and x_j:

$$f_{similar}(x_i, x_j)=\{\min\left(h_c(x_i), h_c(x_j)\right), \min\left(h_t(x_i), h_t(x_j)\right)\} \tag{4.11}$$

The color histogram is the same as the color unary potential computed, and the texture histogram is a SIFT-like feature. The similarty pairwise potential is defined as follow:

$$\phi_{i,j}(y_i, y_j; x_i, x_j, w_{similar})=\begin{cases} w_{similar}\cdot f_{similar}(x_i, x_j) & if \quad y_i\neq y_j \\ 0 & if \quad y_i=y_j \end{cases} \tag{4.12}$$

The pairwise term penalizes the assigment of different labels to similar neighboring superpixels.

4.4 Inference and Learning

As for the inference of CRF model, because the similarity pairwise potential is a metric, the energy function can be solved by the min-cut/max-flow algorithm[7].

The energy function defined as above has the weights $W=\{w_{objness}, w_{color}, w_{similar}\}$. We learn them from a dataset to optimize the segmentation performance by a max-margin structured framework[5].

5. **Experiments**

We perform the experiments on an extended version of the GrabCut dataset[8]. The dataset contains 151 images, user-defined foreground/background seeds and ground truth segmentations. However, the scale of dataset is not enough for testing. So, we extend the original dataset with 200 images from more challenge Pascal VOC 2007 Object Segmentation dataset[9]. The original dataset doesn't provide the foreground/background seeds, so we manually annotate the seeds in the images which have normal size objects.

we employ overlap rate as a measure to evaluate the segment result over one image:

$$ov(s,g)=\frac{|s\cap g|}{|s\cup g|} \tag{5.1}$$

The overlap rate is defined as a ratio of intersection and union between the segmentation s and ground truth g. And we compute the average overlap rate over all images. In order to evaluate the usefulness of objectness unary potiential, we compare the proposed CRF model with/without objectness measure.

Method	Without Objectness	With Objectness
Avg. overlap rate	0.687	0.796

Table 1 : Comparison on the Average Overlap Rate

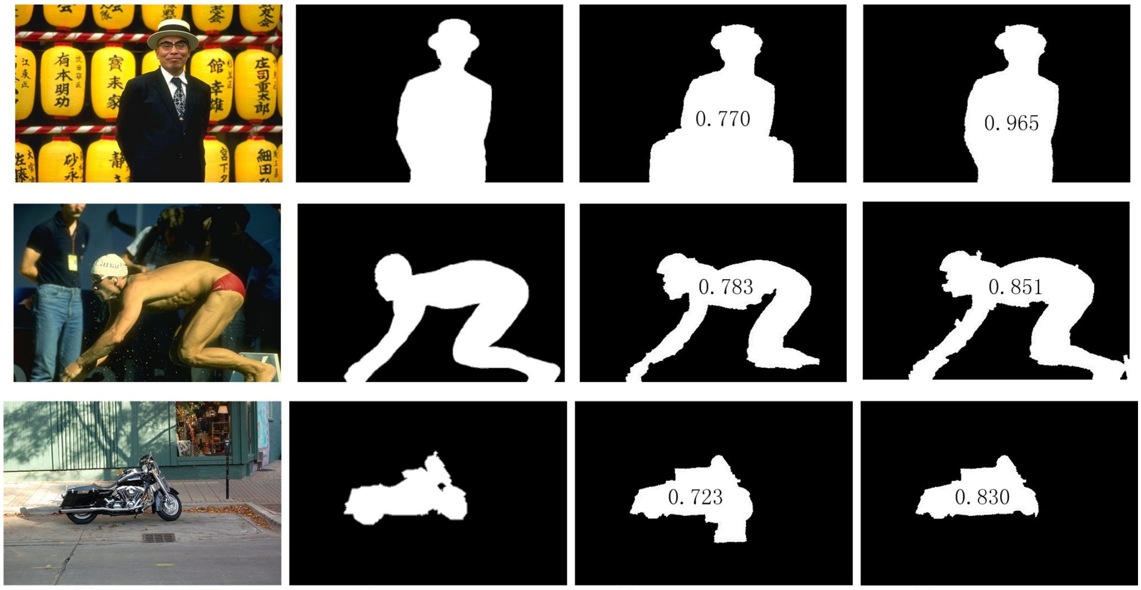
As Table 1 shows, the foreground segment model with objectness measure can reach a higher average overlap rate, which means the objectness providingthe useful object location prior does really help to get a better segmentation. We also provide some segmentation results in Fig. 3.

We also analyze the object-level characteristics' impact on the proposed segmentation model. Specifically, the model is evaluated on five different categories of objects including *Person*, *Cat*, *Horse*, *Car* and *Sheep*. The experimental result is shown in Table 2:

Method	Without Objectness	With Objectness
Person	0.635	0.723
Cat	0.754	0.834
Horse	0.663	0.754
Car	0.697	0.807
Sheep	0.651	0.767

Table 2 : Comparison of Different Categories

As Table 2 shows, different categories have different impacts on the final segmentation result. Among the five categories, all segmentations of *Cat* reach the best performance. Our interpretation for such result is that the category of *Car* tends to have large homogeneous foreground or background or is highly contrasted. However, *Person* objects, often containing large deformation and thin structures, are difficult to segment.



(a). input image (b). ground truth (c). without objectness (d). with objectness

Figure 3: Some Segmentation Results in the Dataset, and the Overlap Rates are Labeled in the Center

6. Conclusion

This paper presents a novel foreground segment model with objectness measure. Because the objectness measure can provide an object location prior, that makes the model especially useful for high accuracy foreground object segmentation. In addition, we also show that the method used in the object detection task can also serve for foreground object segmentation. In the future, we should put forward a better manner to incorporate the objectness measure, and try to find other effective prior information.

References

[1] B. Alexe, T. Deselaers, V. Ferrari. *What is an Object*?[C].Proc IEEE Computer Vision and Pattern Recognition 2010, San Francisco, USA, June, 2010:73-80

[2] E. Rahtu, J. Kannala, M. Blaschko. *Learning a Category Independent Object Detection Cascade*[C].Proc IEEE Computer Vision 2011, Barcelona, Spain, Nov, 2011:1052-1059

[3] C. L. Zitnick, P. Dollár. *Edge Boxes: Locating Object Proposals from Edges*[C]. Proc Springer European Conference Computer Vision 2014, Zurich, Switzerland, Sep, 2014:391-405

[4] R. Girshick, J. Donahue, T. Darrell, J. Malik. *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*[C].Proc IEEE Computer Vision and Pattern Recognition 2014, Columbus, USA, June, 2014:580-587

[5] I. Tsochantaridis, T. Joachims, T. Hofmann, A. Yasemin. *Large Margin Methods for Structured and Interdependent Output Variables*[J]. Journal of Machine Learning Research, 2005, 6(9):1453-1484

[6] R. Achanta, A. Shaji, K . Smith, A. Lucchi, P. Fua, S. Süsstrunk. *SLIC Superpixels: Compared to State-of-the-Art Superpixel Methods*[J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 2012, 34(11):2274-2282

[7] Y. Boykov, V. Kolmogorov. *An Experimental Comparison of Min-cut/Max-flow Algorithms for Energy Minimization in Vision*[J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 2004, 26(9):1124-1137

[8] V. Gulshan, C. Rother, A. Criminisi, A. Blake, A. Zisserman. *Geodesic Star Convexity for Interactive Image Segmentation*[C]. Proc IEEE Computer Vision and Pattern Recognition 2010, San Francisco, USA, June, 2010:3129-3136

[9] M. Everingham, L Van Gool , W. Christopher K. I, W John, Z Andrew. *The pascal visual object classes (voc) challenge*[J]. International journal of computer vision, 2010, 88(2): 303-338.

PoS(ISCC2015)039