Image Segmentation by Cascaded Region Agglomeration
Supplementary material

Zhile Ren
Zhejiang University
jrenzhile@gmail.com

Gregory Shakhnarovich
Toyota Technological Institute at Chicago
greg@ttic.edu

Abstract

In these supplementary materials we provide additional examples of segmentation with our method ISCRA *compares to the results obtained with other methods. We also elaborate on some of the details that could not be fully included in the paper due to space limitations.*

Search for optimal α

Here we describe in more detail the optimization for setting α, the parameters controlling the recall/precision tradeoff in the loss used to train a stage in the ISCRA cascade.

We start with bracketing α between $\alpha_1 = 0$ and some number α_2 large enough to ensure that no merges are made by the learned model; in practice, $\alpha_2 = 600$ is sufficiently high even for the first stage (subsequent stages, as shown in the paper, tend to get lower and lower αs). Since the recall obtained by the stage trained with α is in practice monotonic with α, we can proceed by bracketing α with smaller and smaller interval, until it becomes sufficiently small (we use 0.1 as the minimum interval size to continue binary search).

It is important that the recall is evaluated on a set different from the set used to learn $\mathbf{w}(\alpha)$. Our implementation uses 120 random images (out of 200 BSDS300 training images) as training set (drawn independently in each stage). In the first few stages, when computating the results of merging and evaluation of recall for every α is computationally expensive, we use 30 of the remaining 80 images as tuning set. Beyond these few stages we use the entire 80 images not used for training in the given stage as tuning. Note that since the training/tuning partition is done by independent sampling in each stage, the training and testing sets overlap only partially in subsequent stages, thus making the learning less prone to overfitting. Furthermore, since the algorithm keeps merging regions in each stage, after a few stages the set of regions in a given image changes quite a bit; this also reduces the danger of overfitting.

Algorithm 1: Binary search for α

Given: Training set $\{(\mathcal{R}_i, G_i)\}$, tuning set $\{(\mathcal{R}_j, G_j)\}$, $\rho > 0$
initialize $\alpha_1 = 0$, $\alpha_2 = \langle\text{large number}\rangle$
Starting recall: $r_0 \leftarrow \mathrm{REC}(\{\mathcal{R}_i\})$
while $\alpha_2 - \alpha_1 > \delta$ **do**
  $\alpha \leftarrow (\alpha_1 + \alpha_2)/2$
  compute recall with α on tuning set:
    $r(\alpha) \leftarrow \mathrm{REC}(\mathrm{MERGE}(\{I_i, \mathcal{R}_i, \mathbf{w}^*(\alpha)))$
  if $r(\alpha) > r_0 - \rho$ **then** $\alpha_2 \leftarrow \alpha$ **else** $\alpha_1 \leftarrow \alpha$
Return: α_2

Analysis of system components

All the plots reported in this section are obtained on BSDS300 test set.

As briefly described in the paper, in order to evaluate the importance of different groups of features, we trained versions of ISCRA with reduced feature sets, gradually removing more and and more features. Figure 1 shows the results of this evaluation, in terms of the effect of feature removal on ASA; other measures are affected similarly. Each curve shows ASA as a function of number of segments. E.g., '-SIFT/GC/REG' means removal of all SIFT, geometric context, and region shape features (keeping the color, texton, and OWT-UCM boundary strength features). Each set of features helps to improve performance. We see that the region shape features mostly help for large segments (cyan curve, left part of the plot), SIFT features gradually help more as segment grow (blue curve), adding OWT-UCM boundary feature helps but only a little (since the base system already has the gPb boundary feature). Geometric context is, perhaps surprisingly, helpful mostly for mid-size segments.

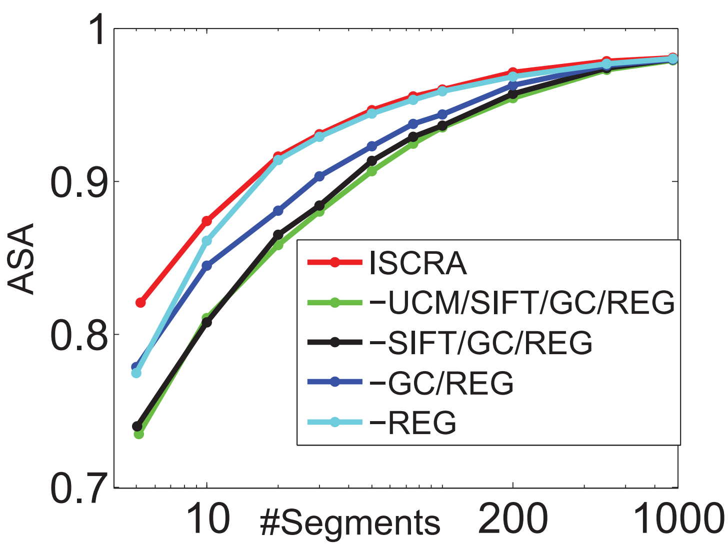


Figure 1. Effect of removal of features (GC: geom. context, REG: region shape features, UCM: OWT-UCM boundary integral) on ASA (BSDS300 test).

Figure 2 shows the ASA curves obtained with ISCRA vs. the non-cascaded system, i.e., the single stage learned from the original superpixels and then applied, without the threshold on Pg, to obtained any desired number of segments. It is clear that its performance deteriorates rapidly once there are fewer than 100 segments left.

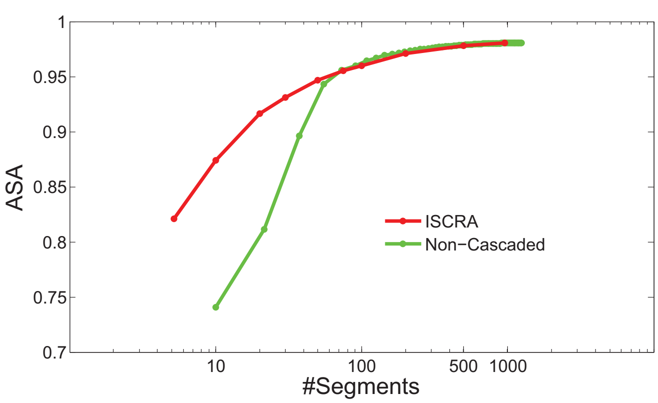
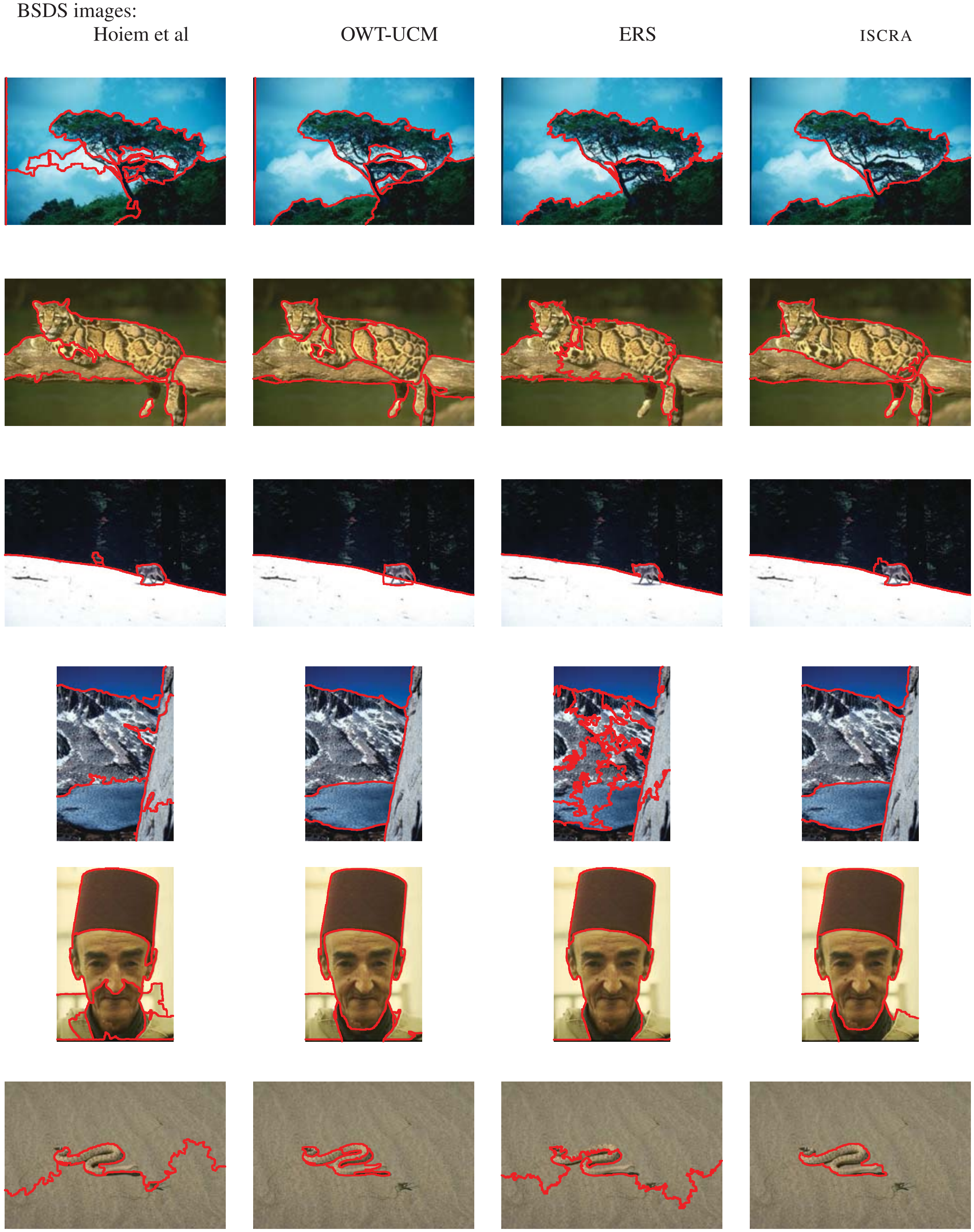


Figure 2. Comparison of ASA with ISCRA (red) to that with non-cascaded system (green).

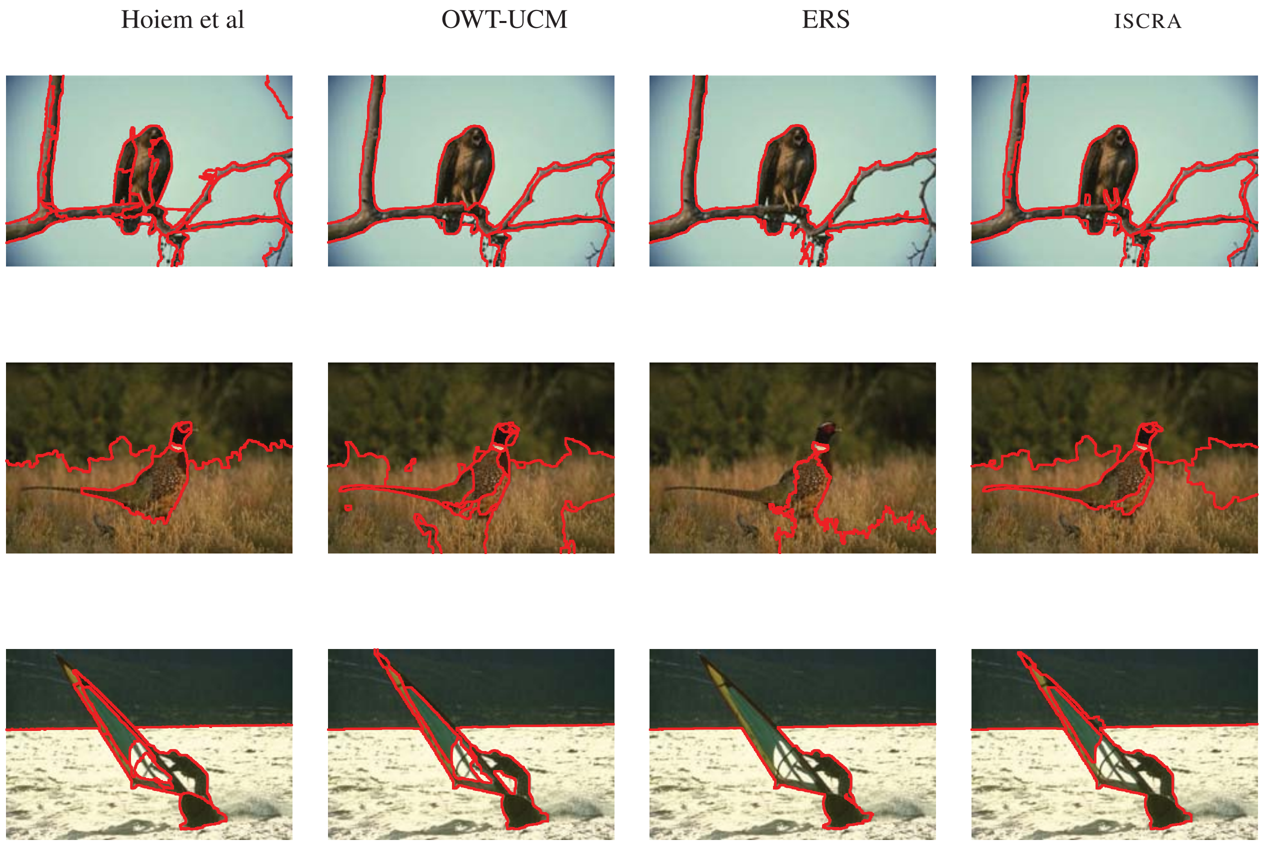
Additional results

In this section we show a few examples of segmentations, from left to right, by: Hoiem et al., OWT-UCM, ERS and ISCRA. For each method, and on each image, we show the segmentation at the scale that maximizes the covering measure for that image (OIS). All images are test images i.e., not used in training ISCRA.

BSDS images:

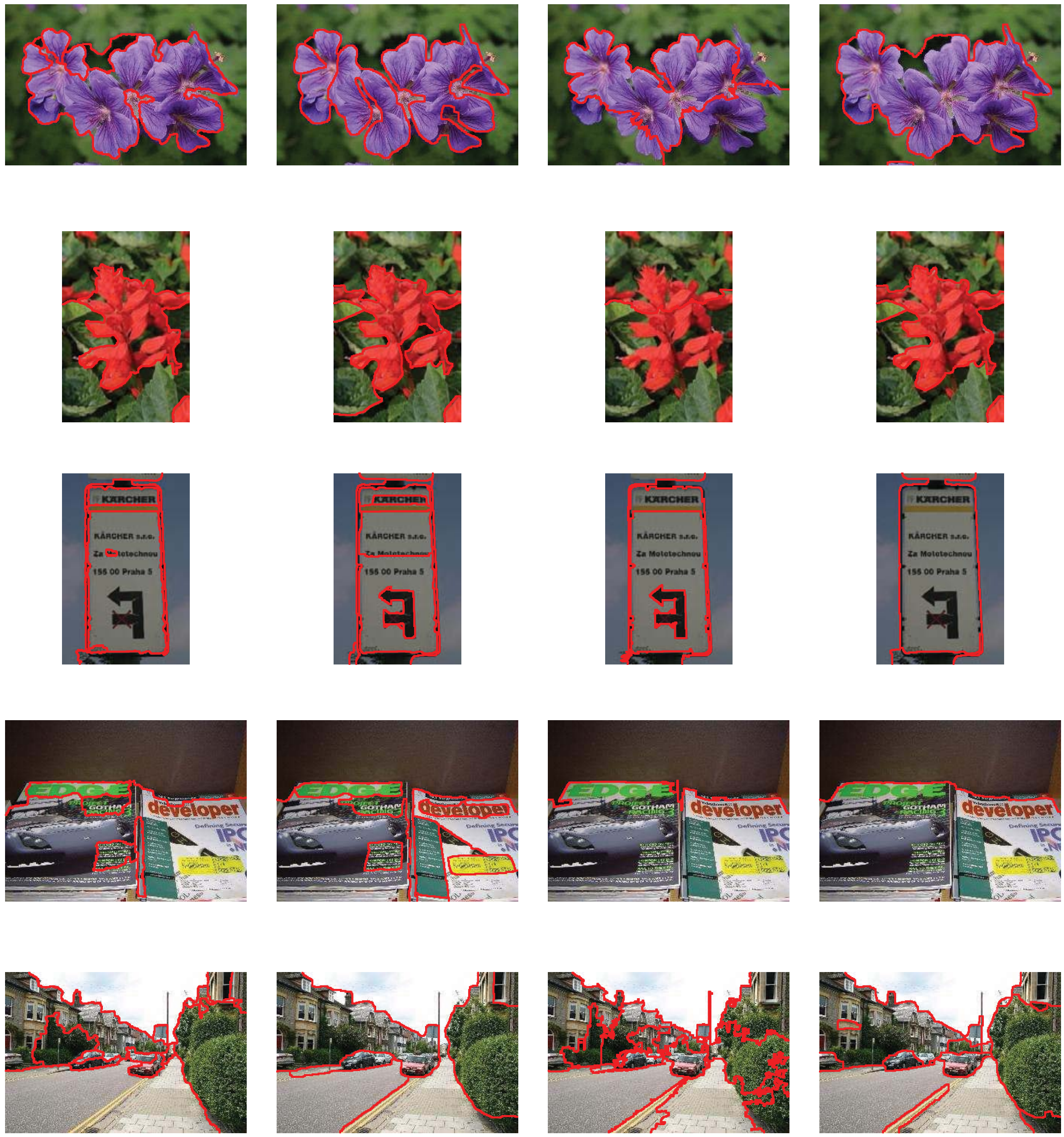


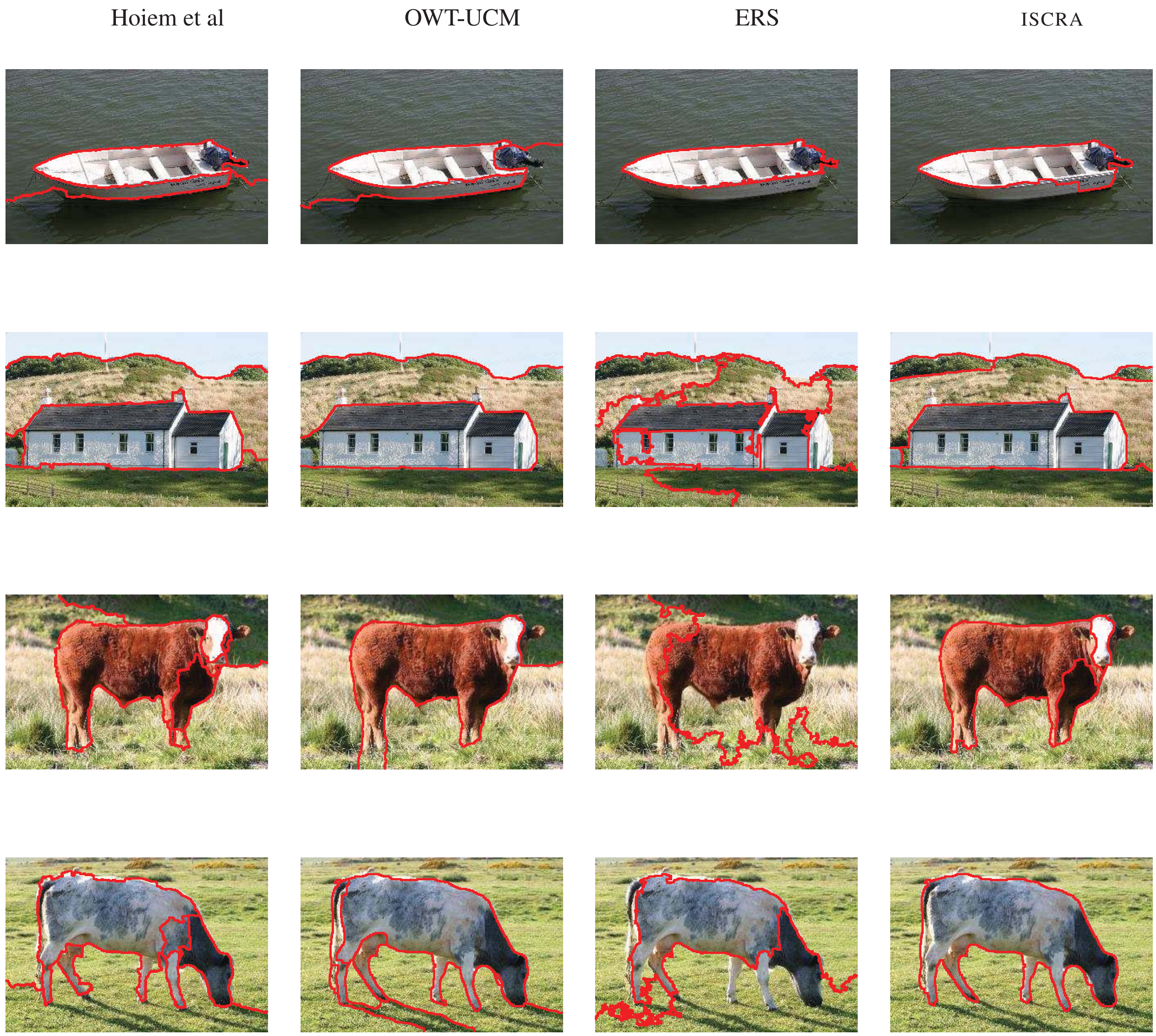
Hoiem et al | OWT-UCM | ERS | ISCRA

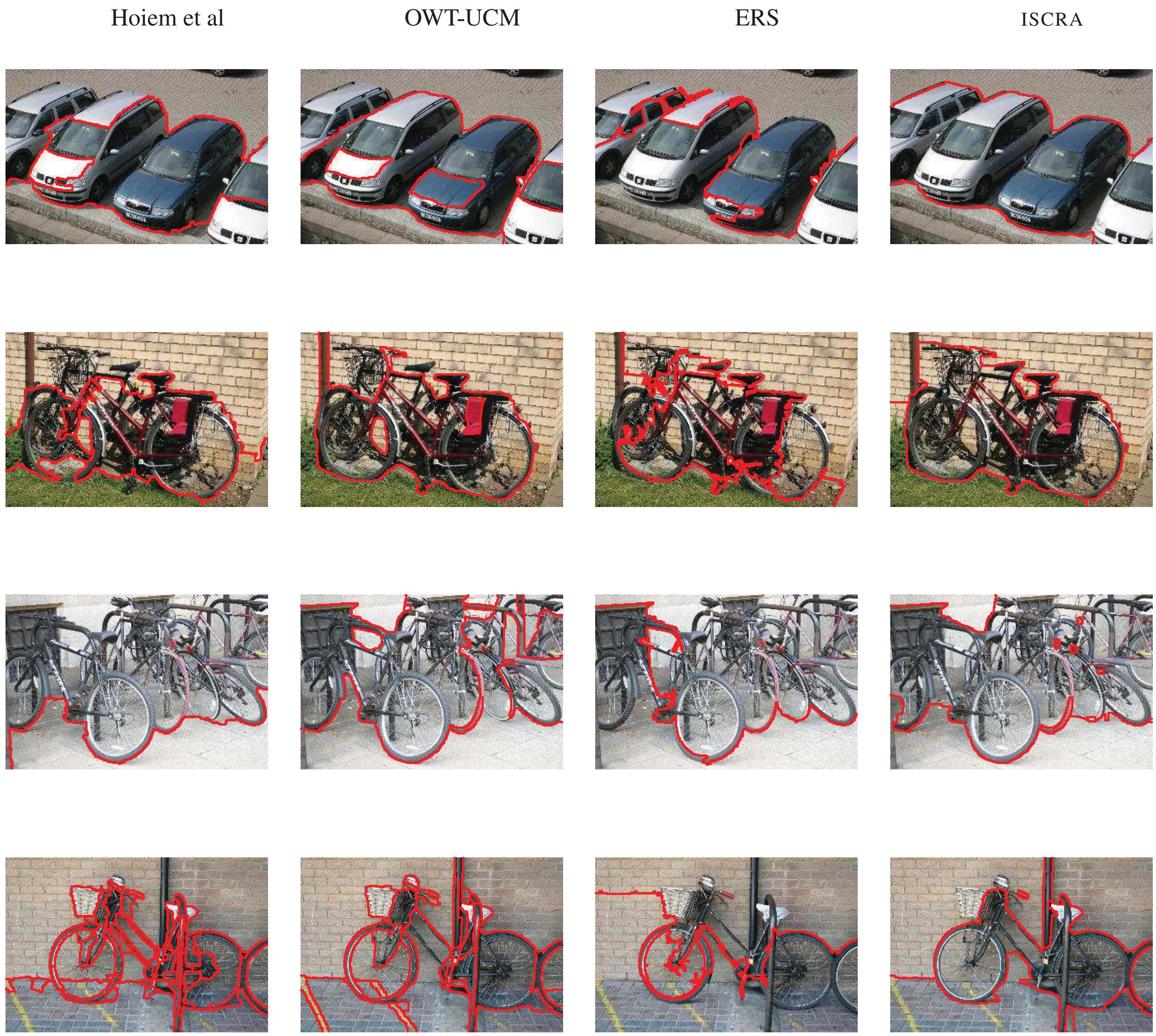

Hoiem et al
OWT-UCM
ERS
ISCRA

MSRC images:

Hoiem et al | OWT-UCM | ERS | ISCRA




Hoiem et al
OWT-UCM
ERS
ISCRA


Hoiem et al
OWT-UCM
ERS
ISCRA

NYU images:

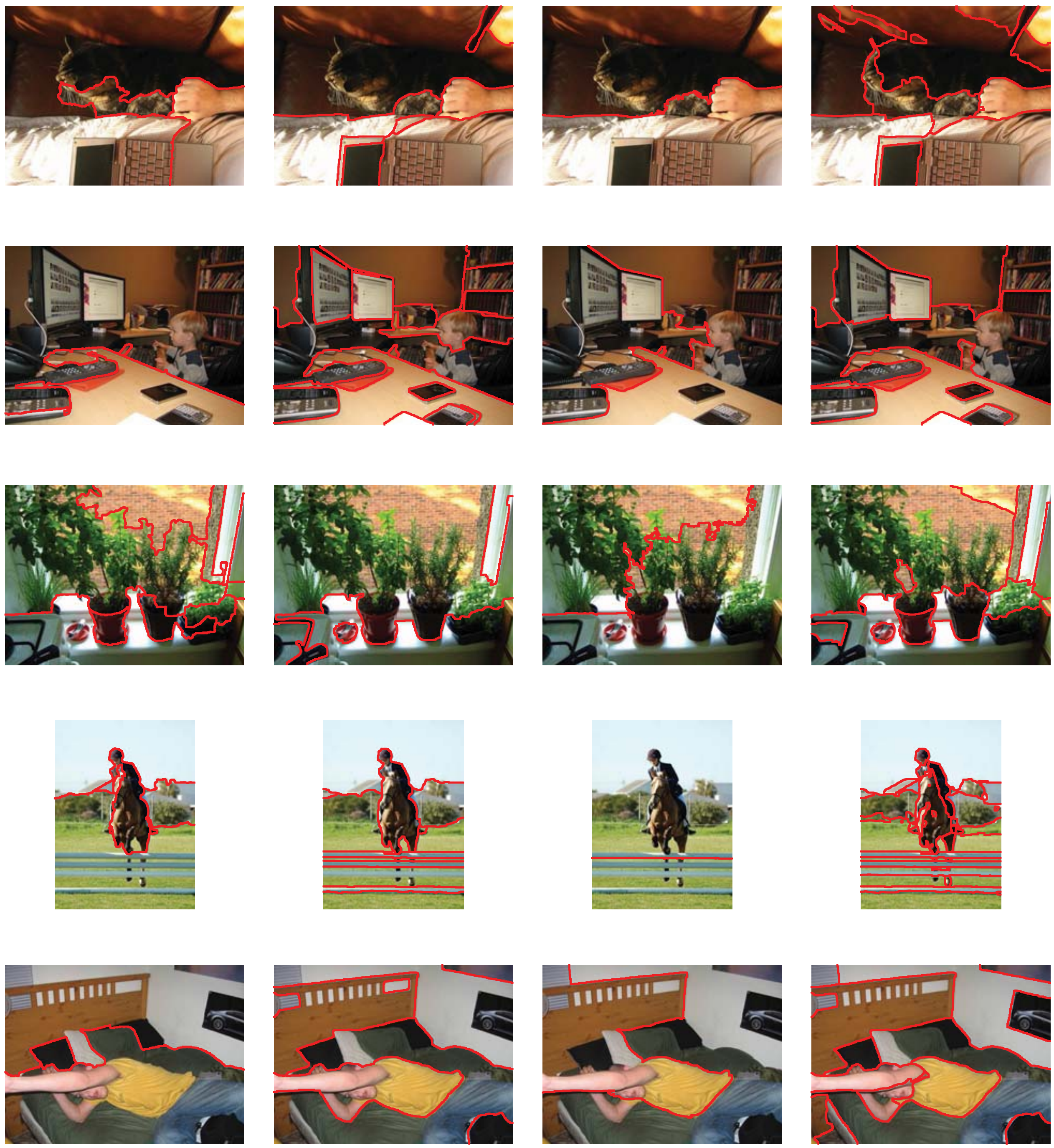
Hoiem et al | OWT-UCM | ERS | ISCRA

SBD images:

Hoiem et al | OWT-UCM | ERS | ISCRA

VOC2012 images

Hoiem et al OWT-UCM ERS ISCRA



Hoiem et al | OWT-UCM | ERS | ISCRA