

Predicting Sufficient Annotation Strength for Interactive Foreground Segmentation - Supplementary Material

Suyog Dutt Jain Kristen Grauman
University of Texas at Austin

suyog@cs.utexas.edu, grauman@cs.utexas.edu

This document contains the following:

1. Example image results where we use the MTurk users' input to initialize the graph cuts interactive segmentation. These examples help illustrate how our method can predict which modality will be sufficiently strong.
2. The user interface used to collect annotations and their timing information on Mechanical Turk.
3. More implementation details about the recognition application result given in Table 1.

1. Example image results from user study

In Figures 1 through 3 we show example images with the user annotations we received from Mechanical Turk users. Figure 1 shows examples where either a bounding box or sloppy contour would be sufficient to achieve good segmentation accuracy. Figure 2 shows examples where a sloppy contour would be sufficient, but a bounding box would not. Figure 3 shows examples where only a tight polygon will suffice to get accurate segmentation.

For all figures, we show the user annotations for all modalities (green annotations) as well as the segmentation that results (red regions) when we use that input with interactive graph cuts. A blue box around an image signifies that our approach predicts that image to be sufficiently handled by the annotation modality specified for that column. So, correct predictions are when the column label where a blue-outlined image appears agrees with the figure label at the top. When the foreground is compact and/or distinctive from the probable background region, our method can correctly predict that a bounding box is enough (Fig. 1). On the other hand, when foreground-background ambiguity is high or the foreground shape complex, it correctly predicts that stronger inputs are needed (Figs. 2 and 3).

In each figure, the last two rows represent failure cases for our method. Failure cases usually occur when salient region detection fails drastically or when our features do not capture the complexity of the foreground object successfully. In some cases, high color similarity between foreground and background also leads to an overestimate in segmentation difficulty.

Overall, we see that in most cases our method correctly predicts the least expensive annotation modality that will succeed. These results with 101 total real users and images from three datasets offer strong support for our idea: requesting the easiest user interaction that is sufficiently strong helps reduce annotation time, while still providing accurate interactive foreground segmentations.

2. User Study Interface

Figure 4 shows the web interface which was used to collect user annotations and timing information. The mechanical turk users were given specific instructions for each annotation modality along with some example annotations. They were also shown the actual dataset ground-truth as a quality control measure.

Bounding Box or Sloppy Contour sufficient

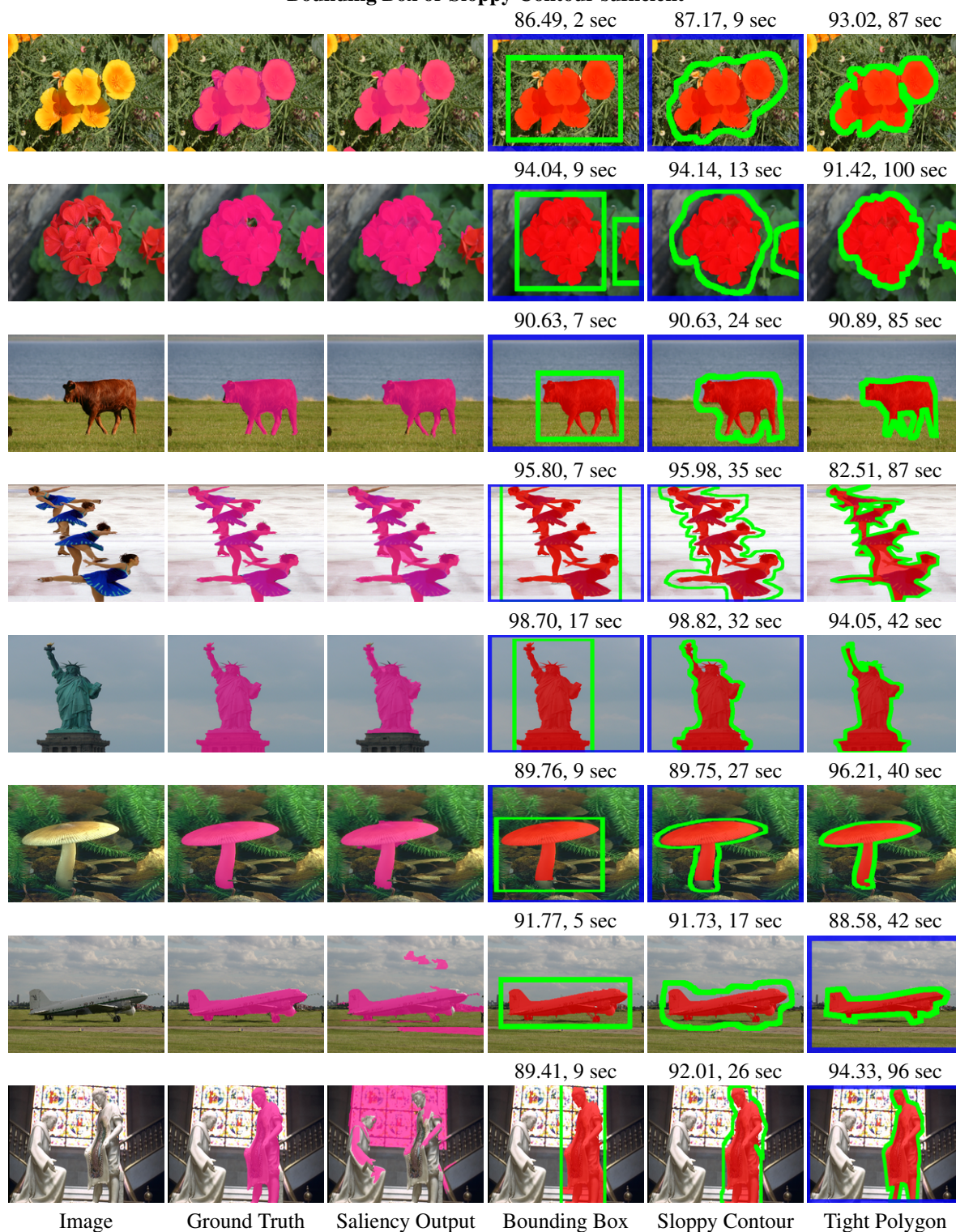


Figure 1: Interactive segmentation results using real MTurk user annotations, where either a bounding box or sloppy contour is sufficient. The ground truth foreground and saliency detector output are shown in pink in second and third column respectively. User annotations are marked in green, and the graph cut segmentation results using those annotations are shown in red. Overlap with ground truth and total annotation time are marked above each image. The images outlined in blue denote the modalities our method predicts to be sufficiently strong. Last two rows are failure cases. Best viewed in color.



Figure 2: Interactive segmentation results using real MTurk user annotations, where a sloppy contour is sufficient but a bounding box is not. The ground truth foreground and saliency detector output are shown in pink in second and third column respectively. User annotations are marked in green, and the graph cut segmentation results using those annotations are shown in red. Overlap with ground truth and total annotation time are marked above each image. The images outlined in blue denote the modalities our method predicts to be sufficiently strong. Last two rows are failure cases. Best viewed in color.

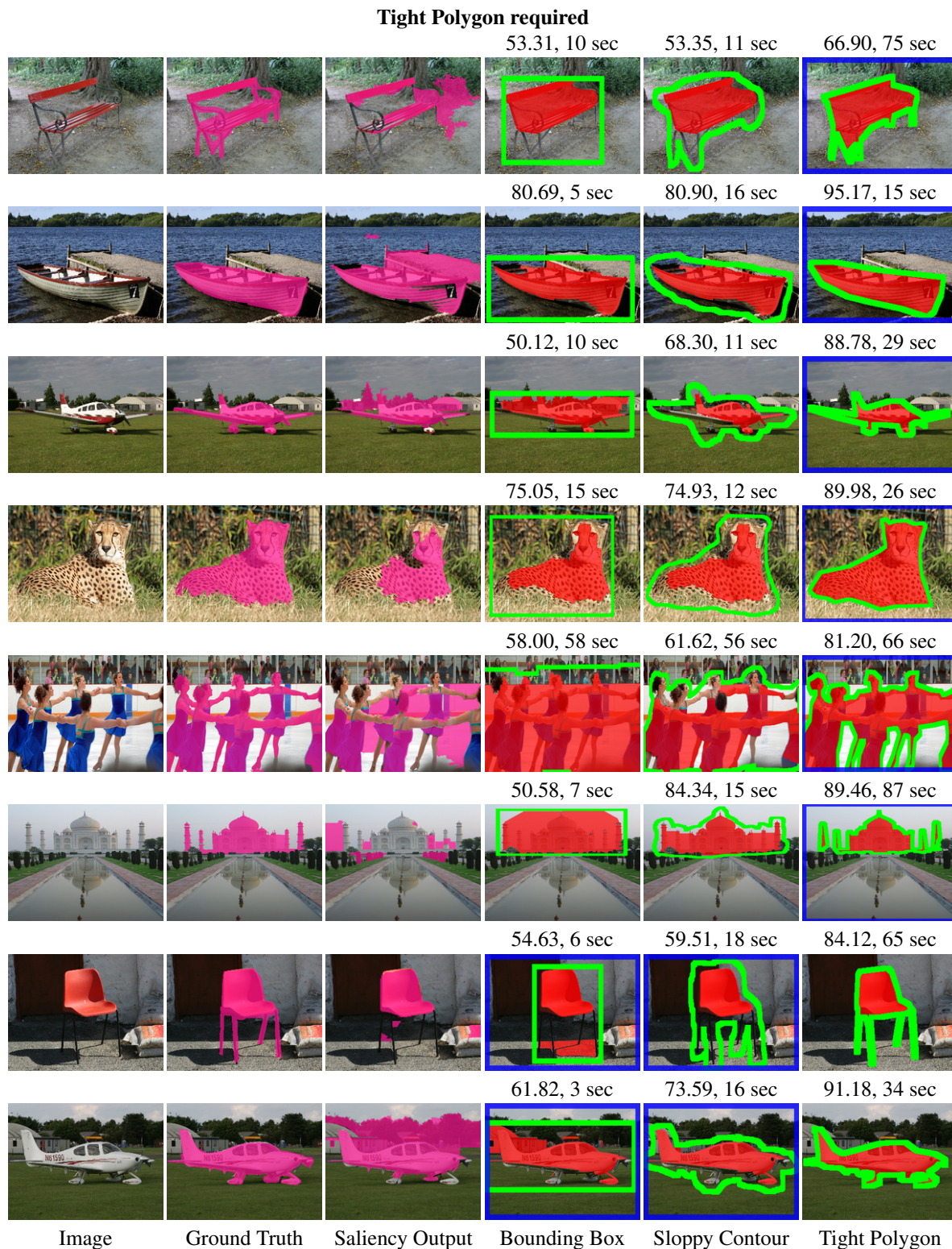


Figure 3: Interactive segmentation results using real MTurk user annotations, where a tight polygon is required. The ground truth foreground and saliency detector output are shown in pink in second and third column respectively. User annotations are marked in green, and the graph cut segmentation results using those annotations are shown in red. Overlap with ground truth and total annotation time are marked above each image. The images outlined in blue denote the modalities our method predicts to be sufficiently strong. Last two rows are failure cases. Best viewed in color.




Bounding Box	Sloppy Contour
<p>Draw rectangle around main objects in the image shown in the red box</p> <p>The main objects are highlighted with green color in the image above the red box</p> <p>Read instructions below (Accept the HIT before starting)</p>  <p>Reset Next Submit Results</p>	<p>Draw rough outlines around main objects in the image shown in the red box</p> <p>The main objects are highlighted with green color in the image above the red box</p> <p>Read instructions below (Accept the HIT before starting)</p>  <p>Reset Next Submit Results</p>
Tight Polygon	Instructions (as shown to annotators)
<p>Click on the object boundaries to very accurately outline the main objects in the image shown in the red box</p> <p>The main objects are highlighted with green color in the image above the red box</p> <p>Read instructions below (Accept the HIT before starting)</p>  <p>Reset Next Submit Results</p>	<p>Bounding Box:</p> <ul style="list-style-type: none"> - This task requires you to draw tight rectangles around main objects in the image shown in the red box above. - A computer program will evaluate the results, so bad work will automatically be filtered out. - Press the left mouse button to start drawing a rectangle, move/drag your mouse (with left button pressed) till you cover the whole object and then release the button. - Press "Reset" to restart the work, if you make a mistake. - Press "Next" to go the next image, and Press "Submit Results" after the last image to finish the HIT. <p>Sloppy Contour:</p> <ul style="list-style-type: none"> - This task requires you to draw a rough outline around main objects in the image shown in the red box above. - A computer program will evaluate the results, so bad work will automatically be filtered out. - Press the left mouse button to start scribbling, move/drag your mouse (with left button pressed) to draw the outline, release the button when done. - You can use multiple scribbles to cover the objects. - Press "Reset" to restart the work, if you make a mistake. - Press "Next" to go the next image, and Press "Submit Results" after the last image to finish the HIT. <p>Tight Polygon:</p> <ul style="list-style-type: none"> - This task requires you to trace a very accurate outline around main objects in the image shown in the red box above. - A computer program will evaluate the results, so bad work will automatically be filtered out. - Start by clicking somewhere and keep extending the outline by clicking at the next point on the object boundary. - You will have to close the object boundary by clicking on the starting point, before submitting. - Press "Reset" to restart the work, if you make a mistake. - Press "Next" to go the next image, and Press "Submit Results" after the last image to finish the HIT.

Figure 4: User interface for our user study to collect annotations on Amazon Mechanical Turk.

3. Implementation details on recognition experiment

In Table 1 of the main text, we show a result where our method helps streamline training a region classifier for object recognition. Space did not permit all implementation details, so we elaborate in the following.

We are given a set of training images known to contain an object of interest, plus cluttered background. We apply our one-shot cascade approach to select which annotation modality will be sufficiently strong to extract the foreground object in each training image with interactive graph cuts. Meanwhile, the baseline approach simply gets a tight polygon on each of the images. Then, the resulting segmented images from either method are used to train a linear SVM classifier that can predict whether a new image contains the object or not. We evaluate using leave-one-out cross validation per class, and score accuracy by normalized overlap with ground truth.

To run this experiment, we use the MSRC dataset, and select those 6 object classes for which a single foreground object is present. For each class, we have 15-30 training images. We extract dense SIFT features sampled on a regular grid with 30 pixels spacing, and cluster into 20 visual words. We divide each image into regions using [1], and represent each region with a histogram of visual words. Each region in a training image is assigned a label based on either the interactive segmentation result predicted to be sufficient (for our method) or the tight polygon ground truth only (for the baseline). We train SVMs with the histograms from the resulting foreground object regions in the training examples. At test time, we classify each region in the test image to localize the object.

Table 1 in the main text shows how accurate recognition is for our method and the baseline, compared to how much annotator time was spent to prepare the training images. Note that when an interactive segmentation fails (i.e., if our method chooses an annotation strength that is insufficient), it will introduce noise into the training set, and therefore can decrease test accuracy. In contrast, if we select the right annotation strength, the training set will be clean, yet we will save human effort.

References

[1] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, Sept. 2004.