MILCut: A Sweeping Line Multiple Instance Learning Paradigm for Interactive Image Segmentation

Jiajun Wu[1]*, Yibiao Zhao[2,3]*, Jun-Yan Zhu[4], Siwei Luo[2], Zhuowen Tu[5]

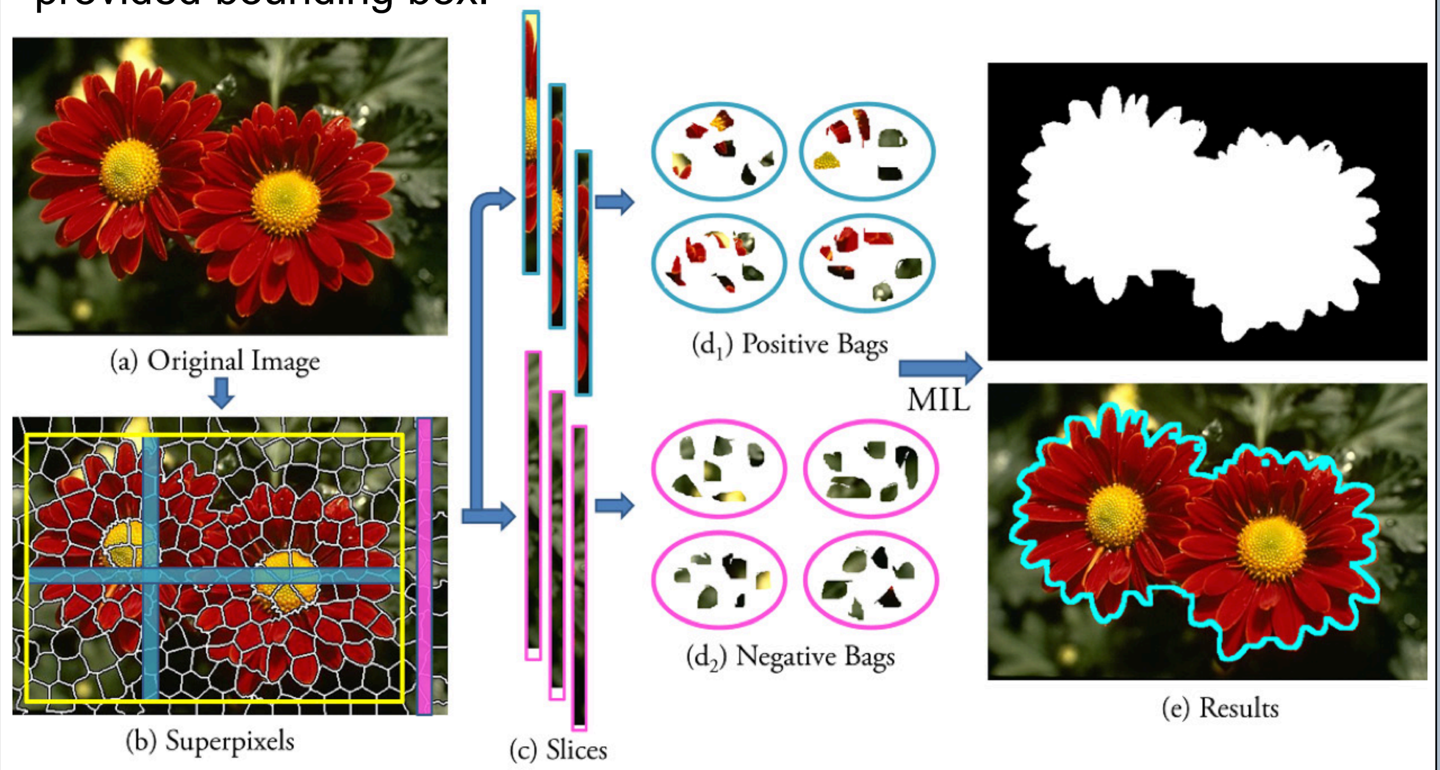
[1]ITCS, Institute for Interdisciplinary Information Sciences, Tsinghua University [2]Beijing Jiaotong University

[3]Department of Statistics, UCLA [4]Computer Science Division, UC Berkeley [5]Department of Cognitive Science, UCSD

CVPR 2014

Motivation

This paper proposes a novel **Multiple Instance Learning** paradigm, namely MILCUT, to segment the foreground object inside a user-provided bounding box.



Multiple Instance Learning

The Multiple Instance Learning (MIL) is a general formulation for dealing with the hidden class labels in noisy input. In MIL, instances (data samples) appear in the form of positive and negative bags:

- **Positive bag**: there is at least one positive instance;
- **Negative bag**: all instances are negatives.

Overview of the Sweeping-Line Strategy

For the interactive segmentation problem, an unknown object of interest is supposed to appear within the bounding box at an unknown location; we also know that image pixels outside the bounding box are background. Therefore, we view the image within the bounding box as "noisy input", and our task is to discover the object under weak supervision with the data in company of outliers.

we provide a **sweeping-line strategy** to convert the interactive image segmentation into a multiple instance learning problem:

- Each horizontal or vertical sweeping line inside the bounding box (a **Positive bag**) must contain at least one pixel from the foreground object;
- Each horizontal or vertical sweeping line outside the bounding box (a **Negative bag**) does not contain any pixel from the foreground object.

In this way, the sweeping-line strategy naturally corresponds to the multiple instance constraints described above.

Validity and Tightness of a Bounding Box

Definition 1. Validity:
For an image I, a bounding box B is valid if the foreground object O completely lies inside the box.

Definition 2. Tightness:
For an image I, a bounding box B is tight if the foreground object O intersects the left, right, top, and bottom border of the bounding box.

Assuming validity and tightness of the bounding box, we then convert the image segmentation task into a MIL problem by considering the horizontal and vertical slices in the bounding box as positive bags and other slices outside the box as negative bags. Either pixels or superpixels could be used as instances. And we proved the Lemma:

Lemma 1. If a bounding box B is valid and tight and the object O inside the bounding box is connected, then the constructed positive and negative bags satisfy multiple instance constraints.

Multiple Instance Learning Formulation

We formulate the interactive image segmentation problem by a structured prediction model named **MILCut-Struct**.
The log-likelihood function is defined as:

$$\mathcal{L}(h) = \mathcal{L}_1(h) + \lambda \mathcal{L}_2(h)$$

The **appearance likelihood model** distinguishes the foreground pixels or superpixels from the clutter background:

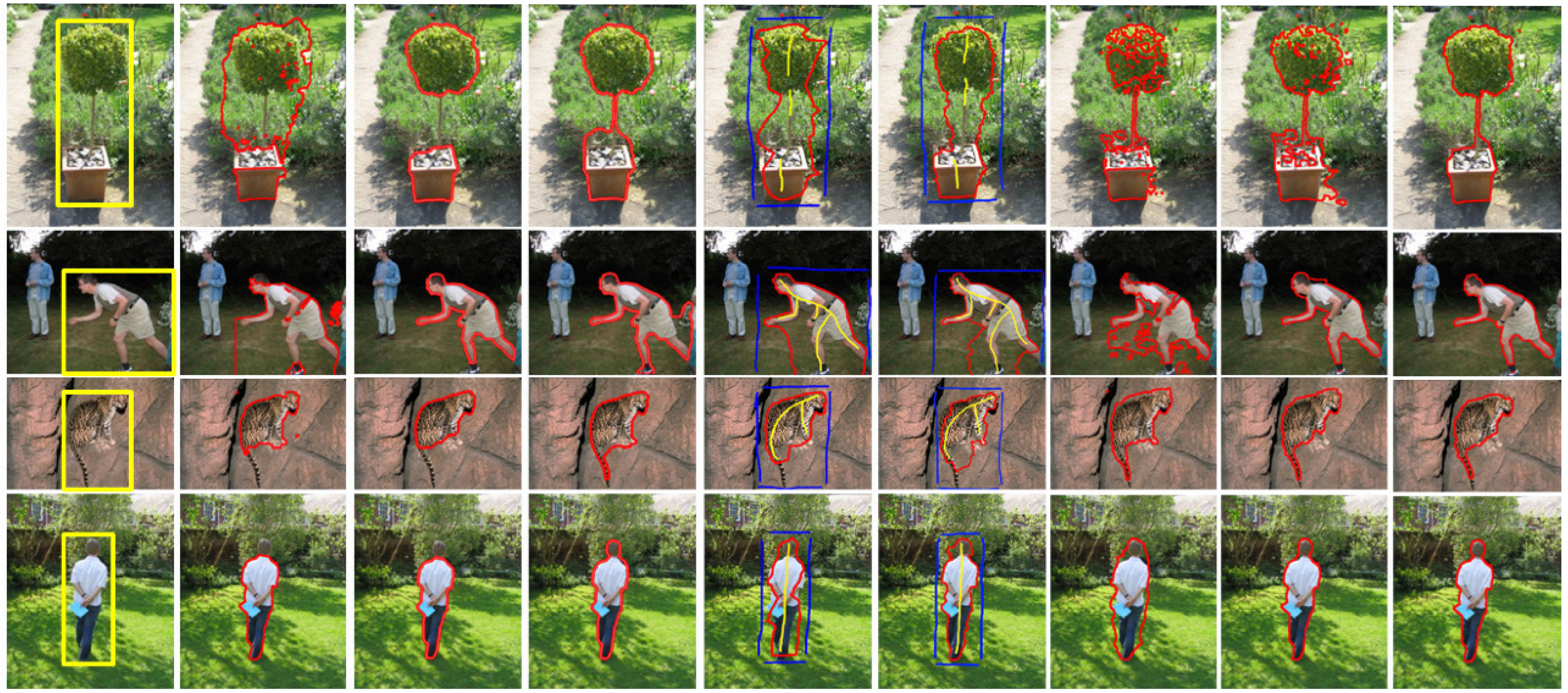
$$\mathcal{L}_1(h) = -\log \prod_i p_i^{y_i} (1-p_i)^{(1-y_i)}$$

The **structural constraints model** enforces the piecewise smoothness in resulting segments:

$$\mathcal{L}_2(h) = \sum_{i=1}^{n} \sum_{(j,k)\in E_i} v_{ijk} \|p_{ij} - p_{ik}\|^2$$

An alternative way of incorporating structural information is applying GraphCut as a post-processing step, which we named **MILCut-Graph.**

Experiment Results on GrabCut dataset



Input bounding box, Initialization method	GrabCut, full bounding box	GrabCut-GC, trimaps	GrabCut-Pinpoint, trimaps	RandomWalker, scribbles	GeodesicSegment, scribbles	MILCut w/o Struct, full bounding box	MILCut-Struct, full bounding box	MILCut-Graph full bounding box
Error rate	8.9	5.9	3.7	5.4	6.8	**6.3**	**4.2**	**3.6**

Experiment Results on Berkeley dataset

Input box,	RegionGrowing,	ObjectExtraction,	MILCut w/o Struct,	MILCut-Struct,	MILCut-Graph
Precision	59	63	**78**	**84**	**83**

Experiments with Noisy Inputs on Weizmann dataset

In general, MILCut explicitly embeds the bounding box prior in the model, and is able to stretch the foreground segment towards all sides of the bounding box. Here shows F-scores on the Weizmann dataset:

Algorithm	ours	[7]	[3]	[17]	[34]	[11]
F-score (%)	**0.89**	0.87	0.86	0.83	0.72	0.57

In real cases, the assumptions we made for the MILCut may not always be satisfied. In this experiment, we consider two distinct situations where multiple instance constraints are not met:

- **Case 1:** The bounding box is not tight. Here shows F-scores on the Weizmann single object dataset with noisy inputs:

	MILCut-Graph	GrabCut [32]
No noise	0.89	0.88
Human noise	0.89	0.86
Machine noise	0.86	0.85

- **Case 2:** The object is not connected. Here shows F-scores on a subset of Weizmann dataset, where each bounding box contains two objects:

Algorithm	ours	[3]	[17]	[11]	[34]
F-score (%)	**0.71**	0.68	0.66	0.61	0.58

Experiments show that MILCut can still obtain better performance than other approaches in these cases.

* The first two authors contributed equally to this work.