Knowledge Leverage from Contours to Bounding Boxes: A Concise Approach to Annotation

Jie-Zhi Cheng, Feng-Ju Chang, Kuang-Jui Hsu, and Yen-Yu Lin

Research Center for Information Technology Innovation, Academia Sinica, Taiwan

1. Summary

- We introduce a concise annotation method to collect training data for class based image segmentation including two steps:
  - Generate multiple tight segments, combining the multiple segment method with the concept of bounding box prior
  - Select the best segment by semi-supervised regression
- Contribution:
  - Present a novel algorithm which integrates the bounding box prior into the concept of multiple image segmentation, and automatically generate multiple tight segments
  - Case the segment selection as a semi-supervised regression problem
  - Demonstrate that our approach provides an effective alternative for manually labeled contours

2. Multiple Tight Segment Generation

- Present an algorithm that automatically generates a set of tight segments for the bounding box of an object, and at least one of these tight segments would approach the object segment

2.1 Tight Segment Via Bounding Box Prior

- First, given a bound box I, partition I into 50 superpixels by means-shift algorithm
- Then, formulate the multiple segment problem as following:

$$\min_{\ell} \sum_{p\in\mathcal{B}} U_p \cdot l_p + \lambda \sum_{(p,q)\in\mathcal{E}} V_{p,q} \cdot |l_p - l_q|$$

subject to $l_p \in \{0,1\}, \forall p$

$$\sum_{p\in C} l_p \geq 1, \forall C \in \Gamma \quad \dashrightarrow \text{ the crossing path constraint}$$

- $\mathcal{E}$: the set of pairs of adjacent superpixels
- U_p: the unary potential
- $V_{p,q}$: the pairwise potential
- Because of the additional constraint, we solve the problem by linear programming, instead of solving it by graph cut

2.2 Multiple Tight Segments

- Generate different seeds by grouping superpixels via different locations and scales
- Unary potential: use the seeds to obtain GMM for background and foreground

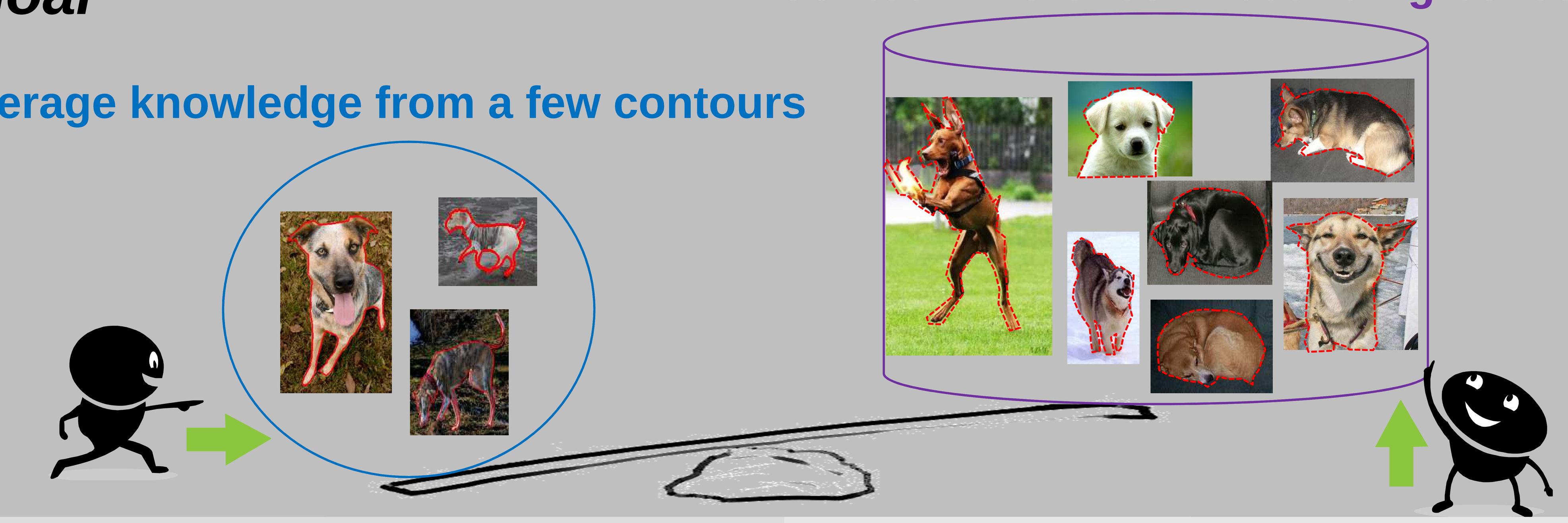
$$U_p = \sum_{u\in p} \log P(c_u|GMM_b) - \log P(c_u|GMM_f)$$

- Pairwise potential:

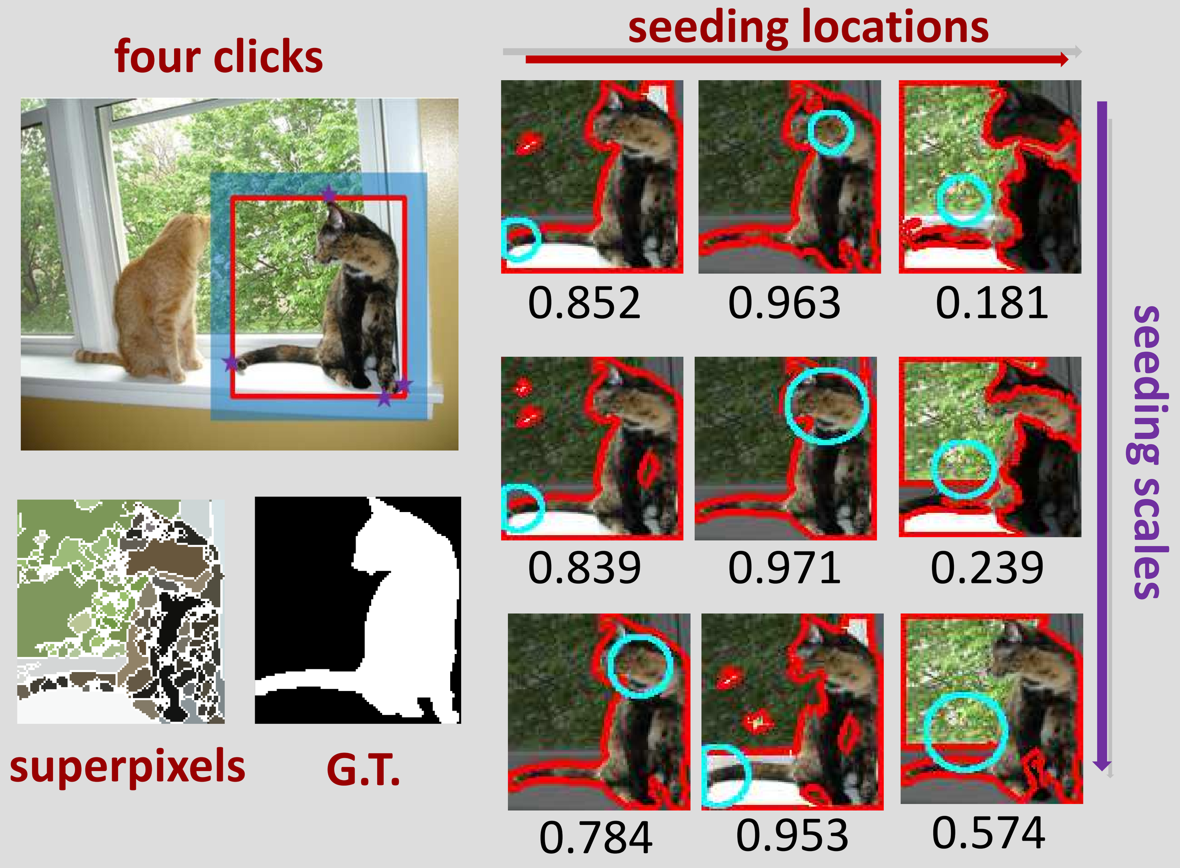
$$V_{p,q} = \sum_{u\in p, v\in q, (u,v)\in N} \frac{1}{dist(u,v)} \cdot \exp(-\beta\|c_u - c_v\|^2)$$

- Use different β, λ, seeds to generate multiple tight segments

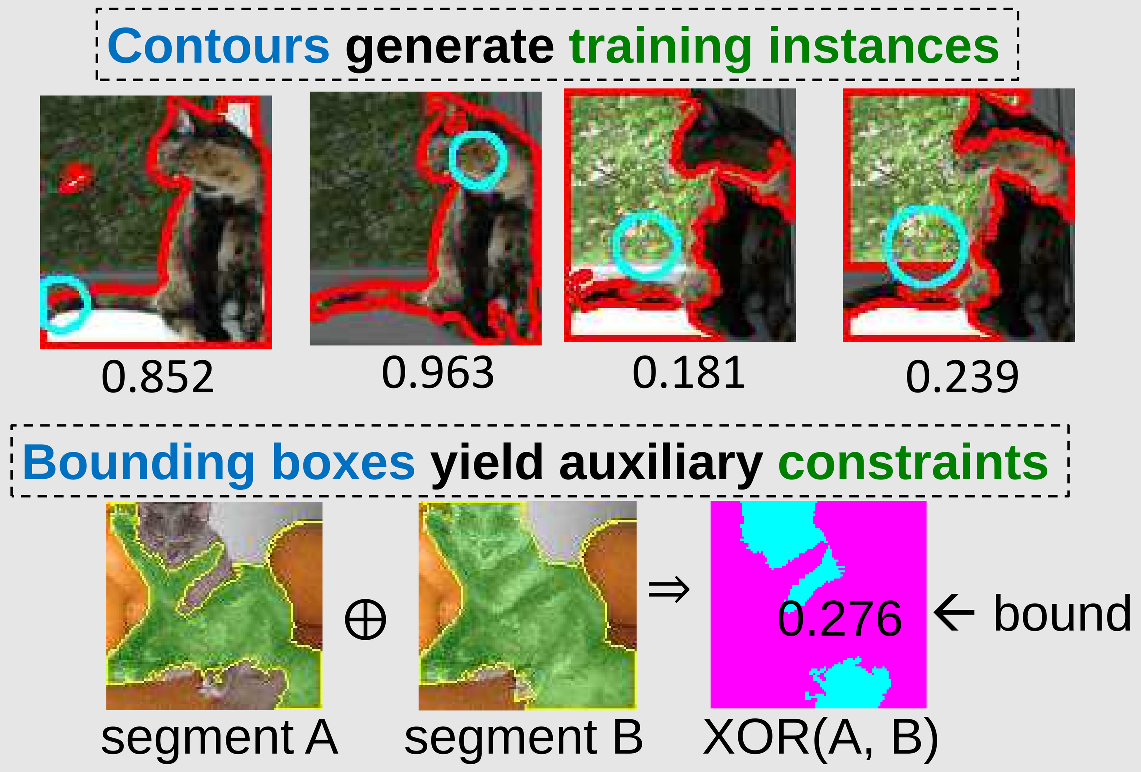
Our Goal



Multiple Tight Segment Generation



Segment Selection by Semi-supervised Regression



3. Segment Selection

- Given a few contours as well as a set of bounding boxes of an object class, we illustrate how to infer the object segments of these bounding boxes by solving a semi-supervised regression problem

3.1 Semi-supervised Regression

- Use the linear regressor to fit the segment selection problem:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$$

  - $\boldsymbol{w}$ and b: parameters of the learned regressor
  - $\boldsymbol{x}$: the segment descriptor
  - Training data: the segments themselves and their accuracy values (calculated by the overlap regions with the ground truth)
  - Semi-supervised constraints: the non-overlap regions between pairs of unlabeled tight segments
- Formulate it as support vector regression with unlabeled constraints, and solve it by efficient tools, e.g., MOSEK

3.2 Segment Descriptor

- Percentage of boundary pixels, boundary edge strength, centroid, major and minor axis lengths, convexity and area
- Foreground and background dissimilarity: color, SIFT, texton

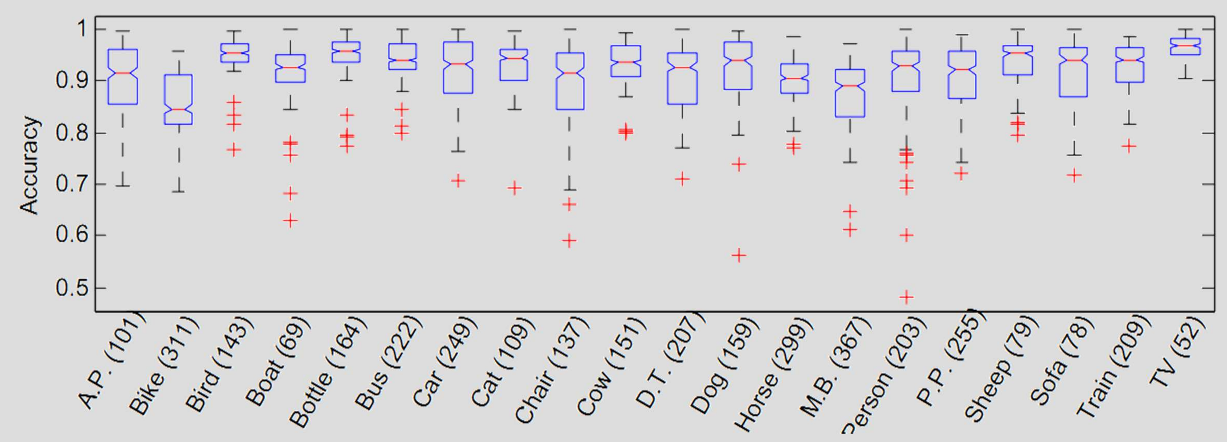
4. Experimental Results

4.1 Dataset: Pascal VOC 2007

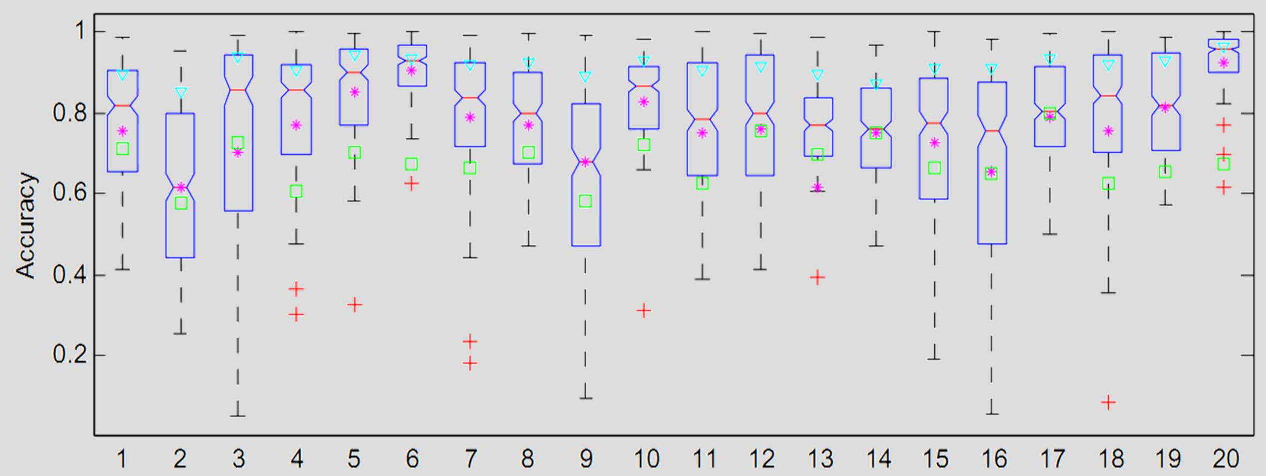
- Contains 21 categories, including 20 object classes with the plus of background
- Each object category contains about 30 to 100 annotated objects, except the class of person, which has more than 300 ones.
- All bounding boxes enclosing the involved objects are resized into the resolution of 80,000 pixels, without changing their aspect ratio

4.2 Exp. I: Multiple Tight Segments

- Verify our assumption, "at least one of these tight segments would approach the object segment"



4.3 Exp. II: Segment Selection



- Show the accuracy rates of the semi-supervised regressor
  - Best tight segments (cyan triangles), GrabCut (green square), Supervised regressor (magenta * signs)

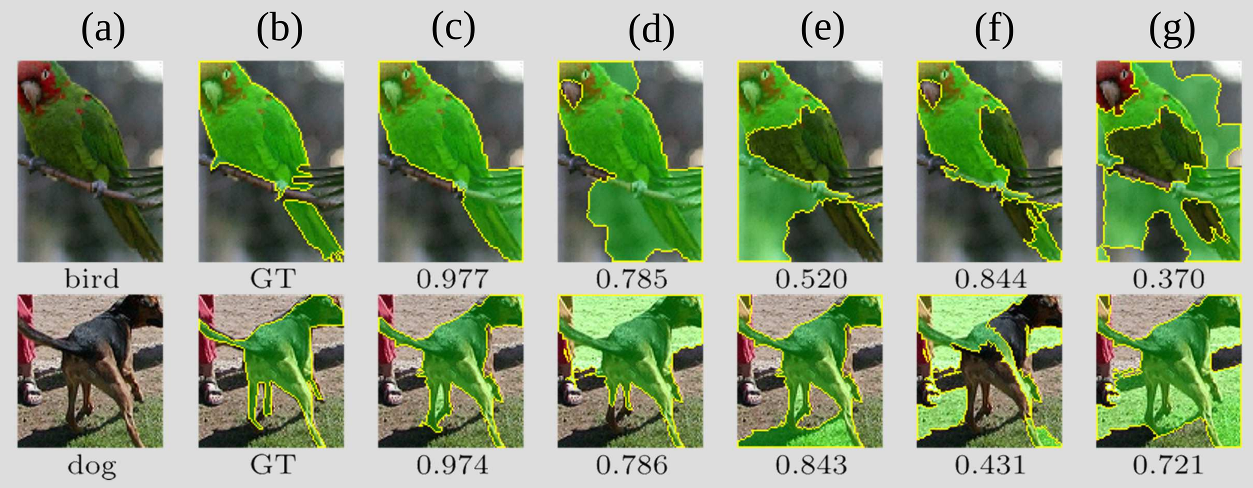
4.4 Exp. III: Image segmentation

- Verify the effectiveness of the concise annotations for the class based image segmentation methods
- Treat the results of segment selection, GrabCut and original ground-truth as training annotations
- Two state-of-the-art class based image segmentation algorithms
  - CRF+N = 0, 2, 4 [Fulkerson, et al., *ICCV* 2009]
  - Hierarchical CRF (HCRF) [Ladick'y, et al., *ICCV* 2009]

	Ground-truth	GrabCut	Supervised Reg.	Semi-Sup. Reg.
HCRF	11.23	9.96	11.06	10.64
CRF+N = 0	14.10	12.47	13.29	13.33
CRF+N = 2	25.26	24.56	26.51	26.51
CRF+N = 4	23.92	21.31	24.81	24.85

4.5 Result Visualization

- Examples of the yielded multiple tight segments: (a) Bounding box. (b) Ground-truth. (c) The best tight segment and its accuracy. (d)~(g) Other tight segments



- Result comparison: (a) Bounding box. (b) Ground-truth. (c) The best tight segment (d) GrabCut (e) supervised regression (f) semi-supervised regression