Discriminative Segment Annotation in Weakly Labeled Video

Kevin Tang[1,2], Rahul Sukthankar[2], Jay Yagnik[2], Li Fei-Fei[1]

[1]Computer Science Department, Stanford University

[2]Google Research

Project page: https://sites.google.com/site/segmentannotation/

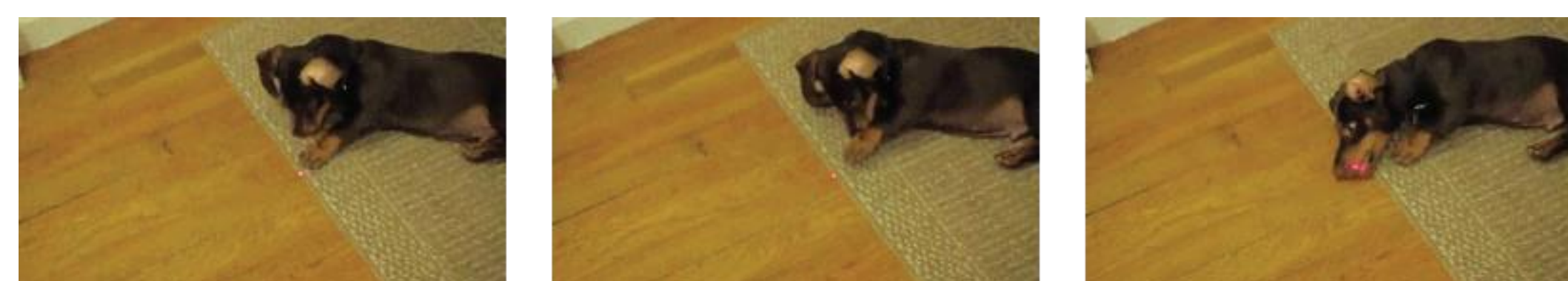## Introduction

- Goal
  - Automatically annotate segments in weakly labeled video taken from YouTube
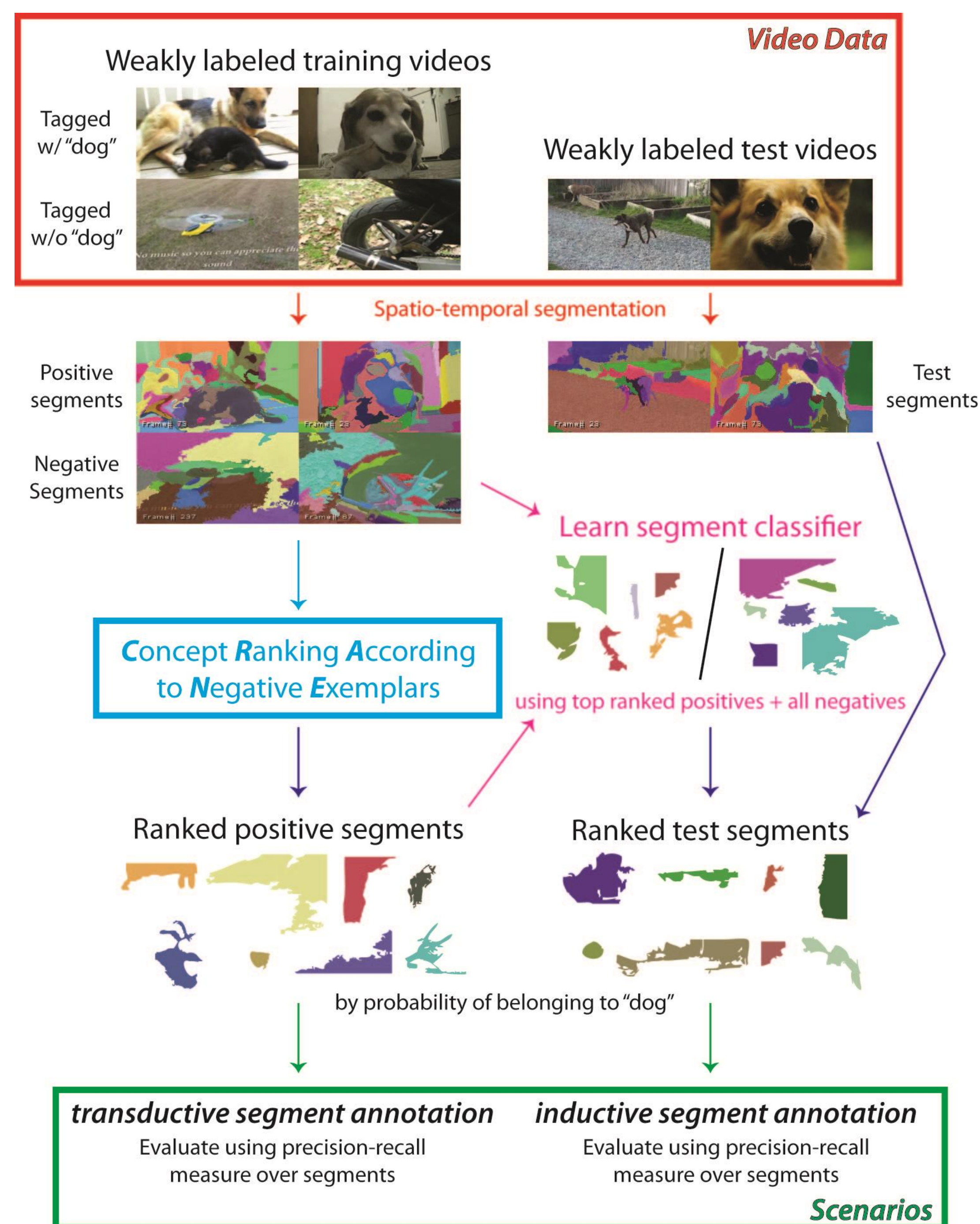

Spatiotemporal segmentation

Semantic object segmentation
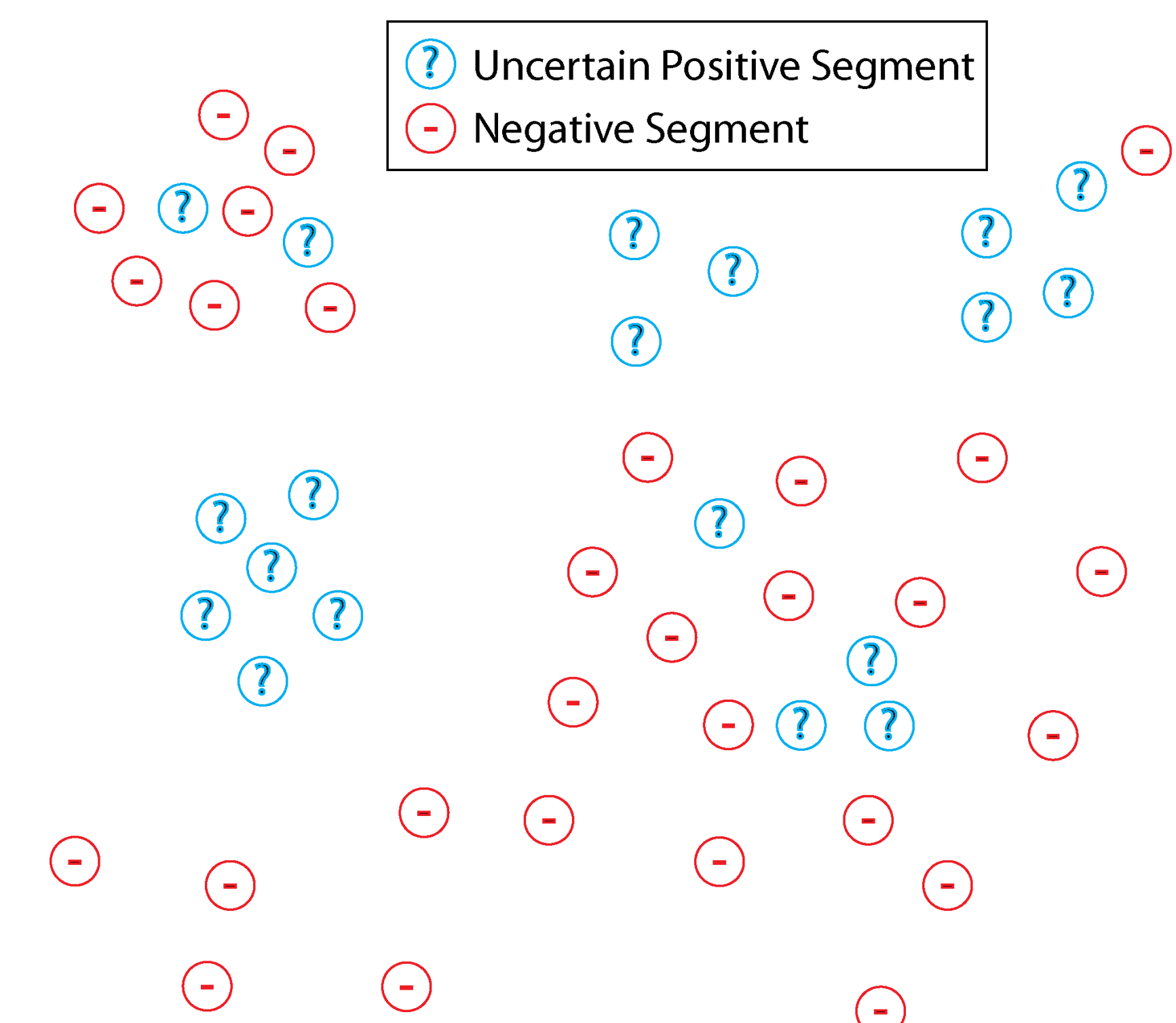
- Challenges
  - Learning from weakly labeled data
  - Handling label noise in YouTube tags
  - Parallelize to deploy over large amounts of YouTube data

Our Problem Setup

Video Data

Weakly labeled training videos

Tagged w/ "dog"

Tagged w/o "dog"

Weakly labeled test videos

Spatio-temporal segmentation

Positive segments

Negative Segments

Test segments

Concept **R**anking **A**ccording to **N**egative **E**xemplars

Learn segment classifier

using top ranked positives + all negatives

Ranked positive segments

Ranked test segments

by probability of belonging to "dog"

transductive segment annotation
Evaluate using precision-recall measure over segments

inductive segment annotation
Evaluate using precision-recall measure over segments
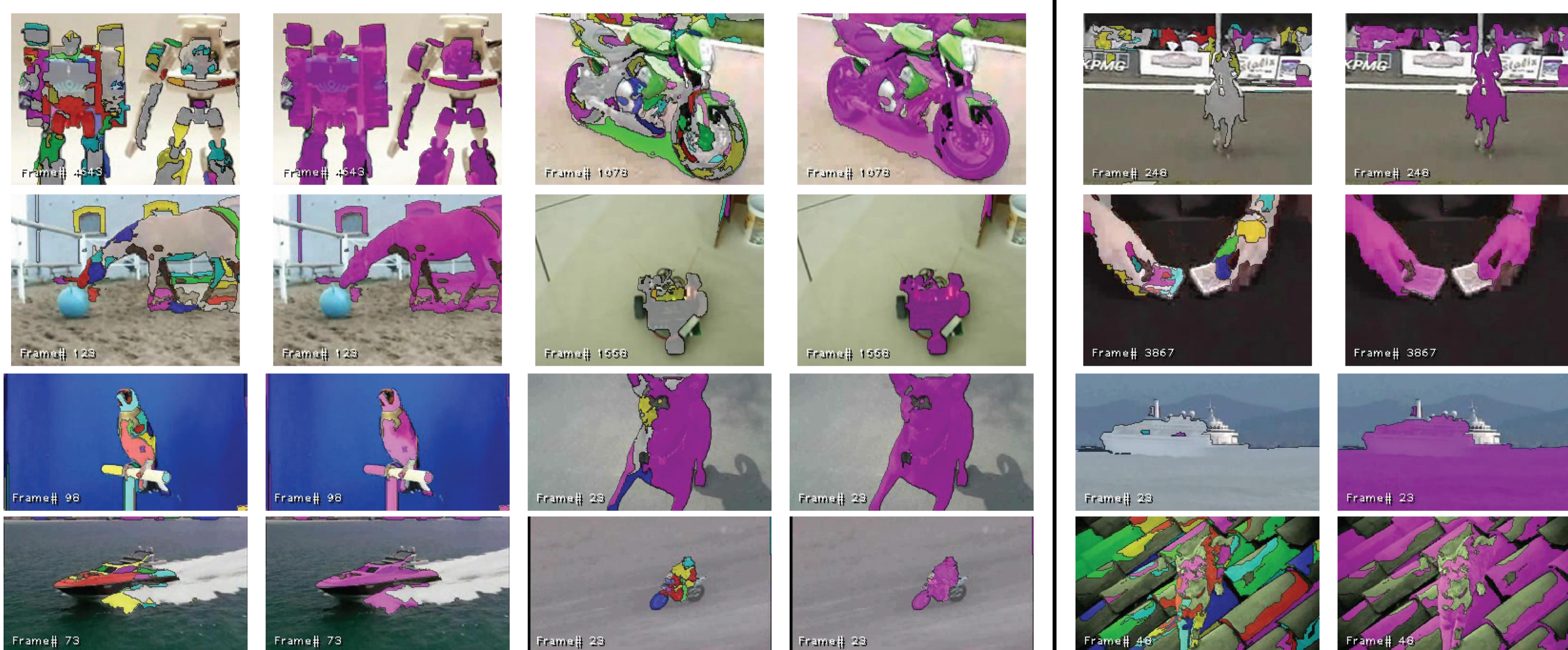
Scenarios

Our Algorithm: CRANE

- Input: uncertain positive segments, large set of negative segments
- Output: ranked positive segments by probability of belonging to our concept

(?) Uncertain Positive Segment
(-) Negative Segment

Intuition: Positive segments are less likely to belong to our concept if they are near many negative segments.

$$S_{\mathrm{CRANE}}(s_i) = -\sum_{z \in \mathcal{N}} \mathbf{1}\left[s_i = \arg\min_{t \in \mathcal{P}} \left(\mathrm{dist}(t, z) \right) \right] \cdot f_{\mathrm{cut}}\left(\mathrm{dist}(s_i, z) \right)$$

Sample Object Segmentations

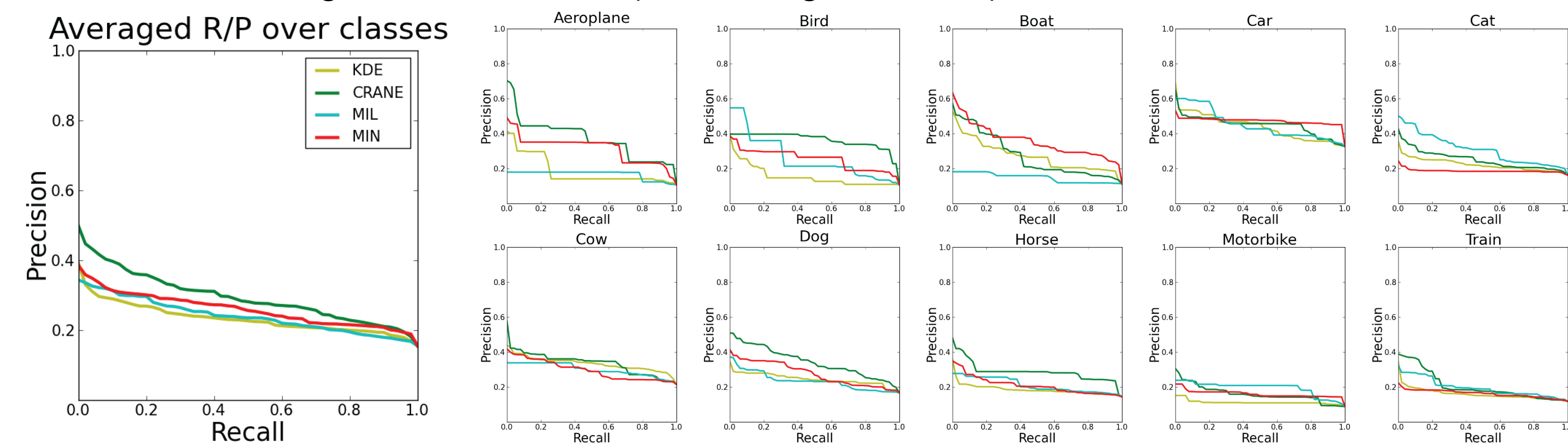

Inductive Segment Annotation [top two rows]
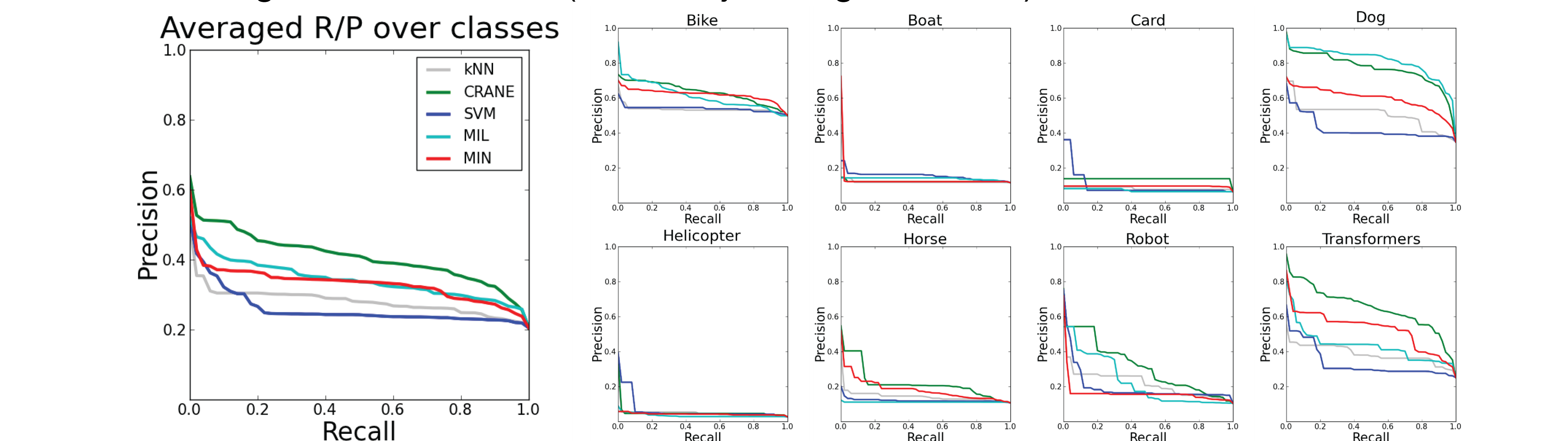
Transductive Segment Annotation [bottom two rows]

Common Failure Cases

[1] P. Siva, C. Russell, and T. Xiang. In defence of negative mining for annotating weakly labelled data. ECCV 2012.

[2] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Efficient hierarchical graph-based video segmentation. CVPR 2010.

[3] G. Hartmann et al. Weakly supervised learning of object segmentations from web-scale video. ECCV 2012 Workshop.

[4] A. Prest et al. Learning object class detectors from weakly annotated video. CVPR 2012.

Quantitative Results

- Transductive Segment Annotation (annotating a dataset)

Averaged R/P over classes

KDE, CRANE, MIL, MIN

Aeroplane, Bird, Boat, Car, Cat, Cow, Dog, Horse, Motorbike, Train

- Inductive Segment Annotation (novel object segmentation)

Averaged R/P over classes

kNN, CRANE, SVM, MIL, MIN

Bike, Boat, Card, Dog, Helicopter, Horse, Robot, Transformers