Co-segmentation of Non-homogeneous Image Sets

ICIP 2018

by

Avik Hati, **Subhasis Chaudhuri**, Rajbabu Velmurugan

Electrical Engineering Department, Indian Institute of Technology Bombay

Image Co-segmentation

- Find co-occurring objects in a set of crowd-sourced images
  - similar feature matching
- Process multiple images simultaneously
  - Common object detected, not recognized






Images from internet

Applications

- Semi-supervised image foreground segmentation
  - Ground-truth annotation

- Image similarity measure

Image courtesy: MSRC dataset

- Object classification

Co-segmentation output

Input image

Objects

Image courtesy: Image pair dataset [Li *et al.*, 2011]

Existing works

- Supervised approaches
  - use scribbles
- Markov random field model based approaches
  - extension of single image segmentation
- Saliency based approaches
  - initialization

Problem definition

- Co-segmentation of a large number (N) of images
- Non-homogeneous image set
  - Common object(s) present only in $M \leq N$ images

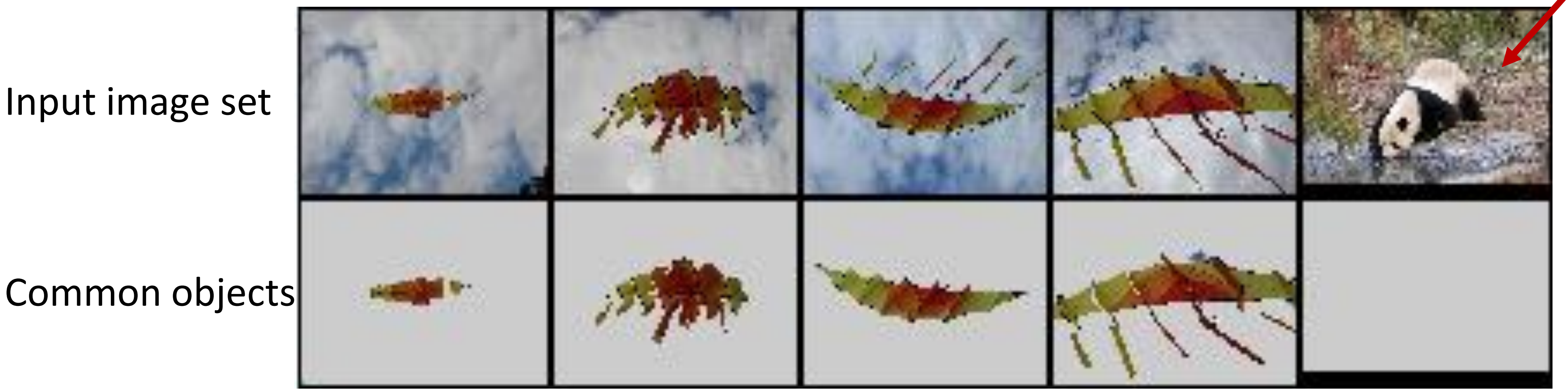


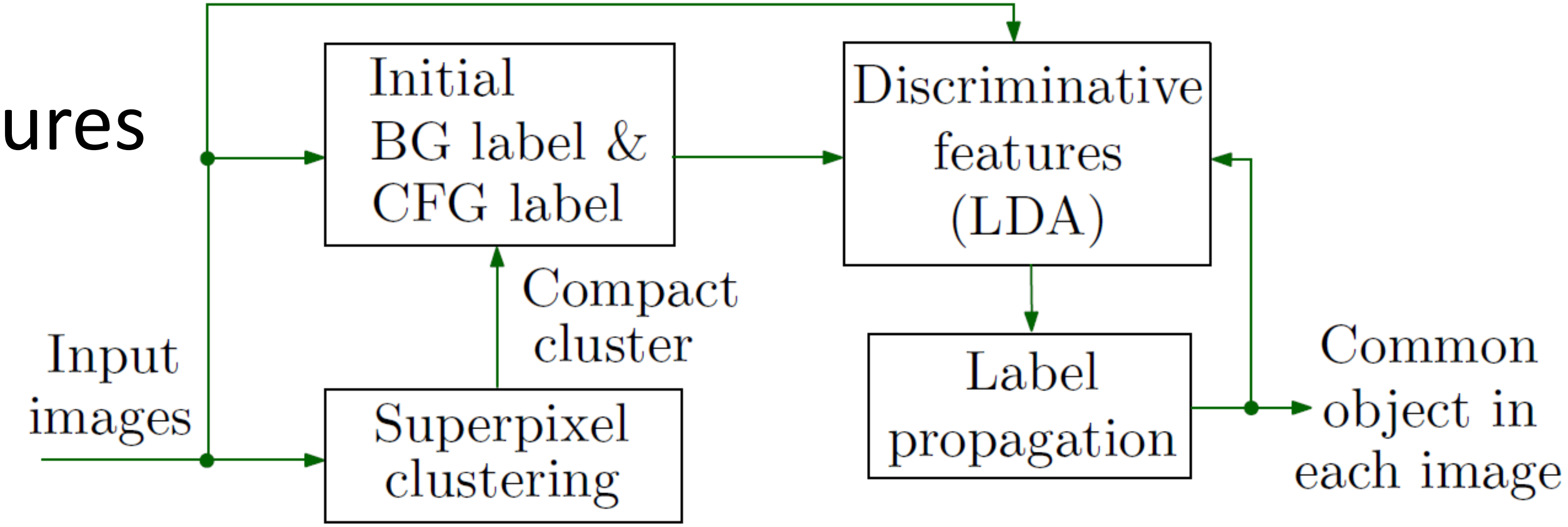
Image courtesy:
iCoseg dataset

Proposed solution

- Low-level features
  - Mean Lab color
  - SIFT
- Mid-level features
  - Bag-of-words
- Discriminative features
  - LDA
- Assign labels

SLIC superpixel segmentation

Image courtesy: Oxford flower dataset



Input images

Initial BG label & CFG label

Compact cluster

Superpixel clustering

Discriminative features (LDA)

Label propagation

Common object in each image

Common foreground seeds

- Superpixel clustering
- Average *spatial compactness* of every cluster-*j*

$$\Gamma_j = \left(\sum_{i=1}^{N} \frac{\sigma_{ij}^2}{A_{ij}} \right)^{-1}$$

SP spatial variance

SP area

  - highest compactness

$$p = \arg\max_j \Gamma_j$$

Cluster of interest

$$C_F = \{\mathbf{s} \in \text{cluster-}p\}$$

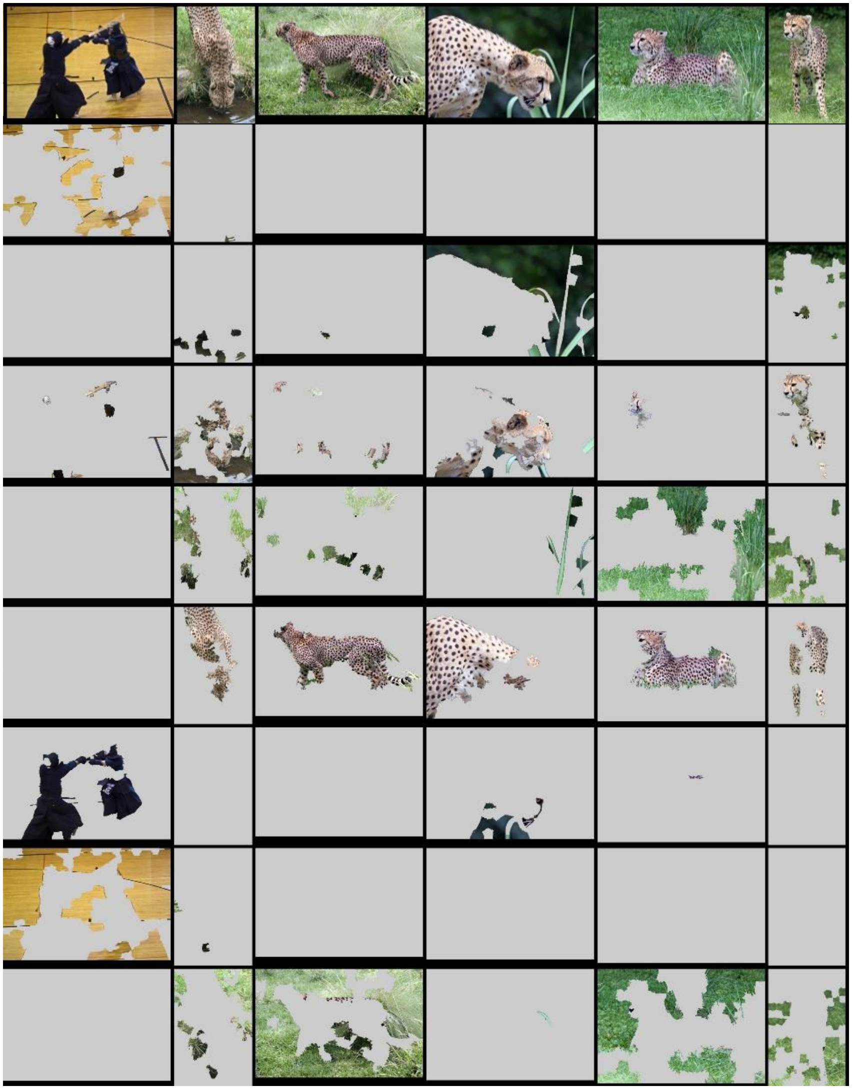


Image courtesy: iCoseg dataset

Background seeds

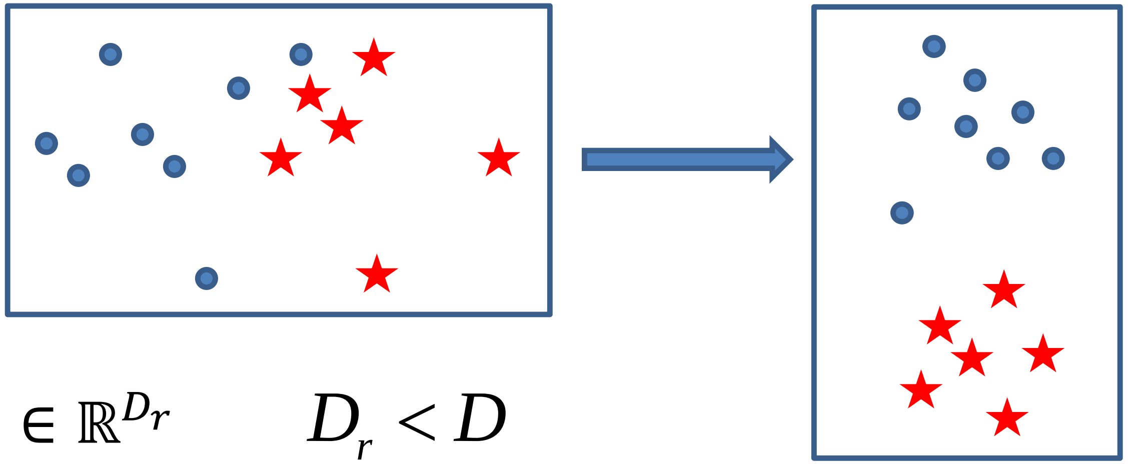
- Background probability of every superpixel [Zhu *et al.*,CVPR 2014]
- Choose superpixels

$$\{\mathbf{s} : P(\mathbf{s}) > 0.99\}$$

  - Less false positives
- Clustering
- Common and different background

W. Zhu, S. Liang, Y. Wei and J. Sun, "Saliency Optimization from Robust Background Detection," *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2814-2821.

Discriminative feature (1/2)



- Feature vector $\mathbf{f} \in \mathbb{R}^{D}$
- Projection $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_{D_r}]$
  - smaller dimension $\mathbf{W}^T\mathbf{f} = \mathbf{f}^{/} \in \mathbb{R}^{D_r}$ $D_r < D$
- Feature points in same class are closer in projected domain
- Use linear discriminant analysis

LDA: Bishop, 2006

Discriminative feature (2/2)

- Between-class variance

$$\mathbf{V}_b = \sum_{i=1}^{K} \frac{n_i}{n_a} \left(\mathbf{m}_i - \overline{\mathbf{m}}\right)\left(\mathbf{m}_i - \overline{\mathbf{m}}\right)^T$$

- Within-class variance $\mathbf{V}_w = \sum_{i=1}^{K} \frac{n_i}{n_a} \mathbf{V}_i$

- $\mathbf{W} =$ Eigenvectors of $\mathbf{V}_w^{-1}\mathbf{V}_b$

K : # of clusters

n_i : class cardinality

n_a : total # of superpixels

$\mathbf{m}_i$: class mean

$\overline{\mathbf{m}}$: all class mean

Label propagation (1/2)

- Seed superpixels labels

$$\begin{array}{ccl} \mathbf{Y} = & \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{n_a \times K} & \begin{array}{l} \longleftarrow \mathbf{s}_1 \in C_{B_2} \\ \mathbf{s}_2 \\ \longleftarrow \mathbf{s}_3 \in C_F \\ \longleftarrow \mathbf{s}_4 \in C_{B_1} \\ \mathbf{s}_5 \end{array} \\ & \begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ C_F & C_{B_1} & C_{B_2} \end{array} & \end{array}$$

Seed superpixels colored

- Find Label matrix $\mathbf{L}$ containing label of all superpixels

similarity matrix

$\mathbf{D}$: Diagonal matrix

- Minimize $\mathbf{L}$

$$\sum_{i,j=1}^{n_a} \mathbf{S}_{ij} \left\| \frac{1}{\sqrt{\mathbf{D}_{ii}}} \mathbf{L}_i - \frac{1}{\sqrt{\mathbf{D}_{jj}}} \mathbf{L}_j \right\|^2 + \alpha \sum_{i=1}^{n_a} \left\| \mathbf{L}_i - \mathbf{Y}_i \right\|^2$$

D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. "Learning with local and global consistency." In *Advances in neural information processing systems*, pp. 321-328. 2004.

Label propagation (1/2)

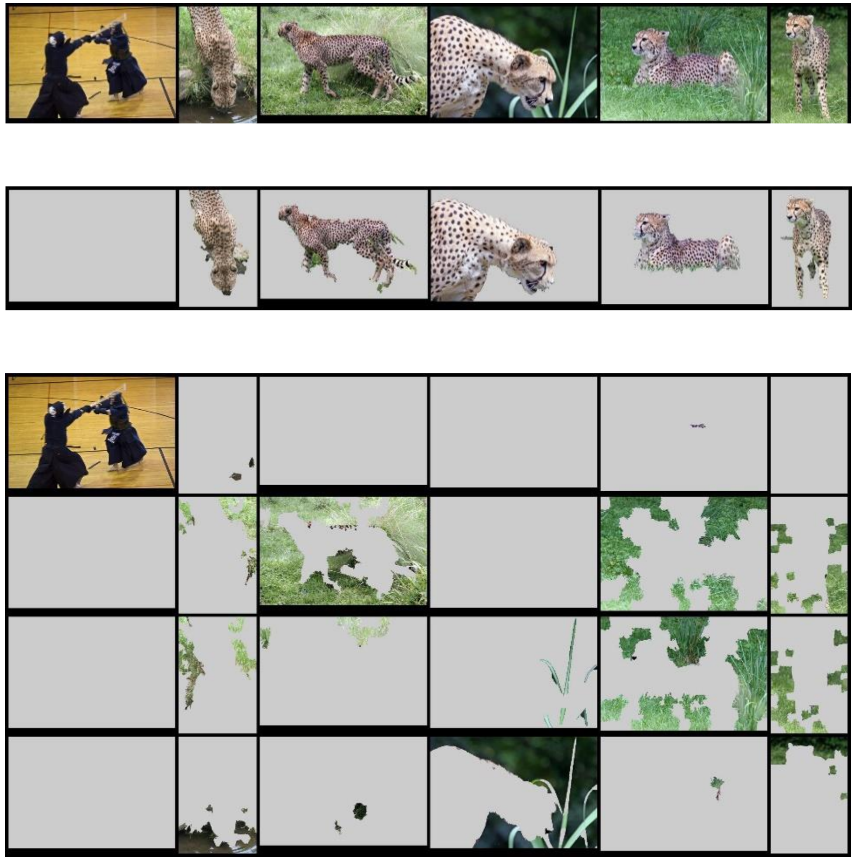
- Solution

$$\mathbf{L}^* = \beta_1(\mathbf{I} - \beta_2\mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2})^{-1}\mathbf{Y}$$

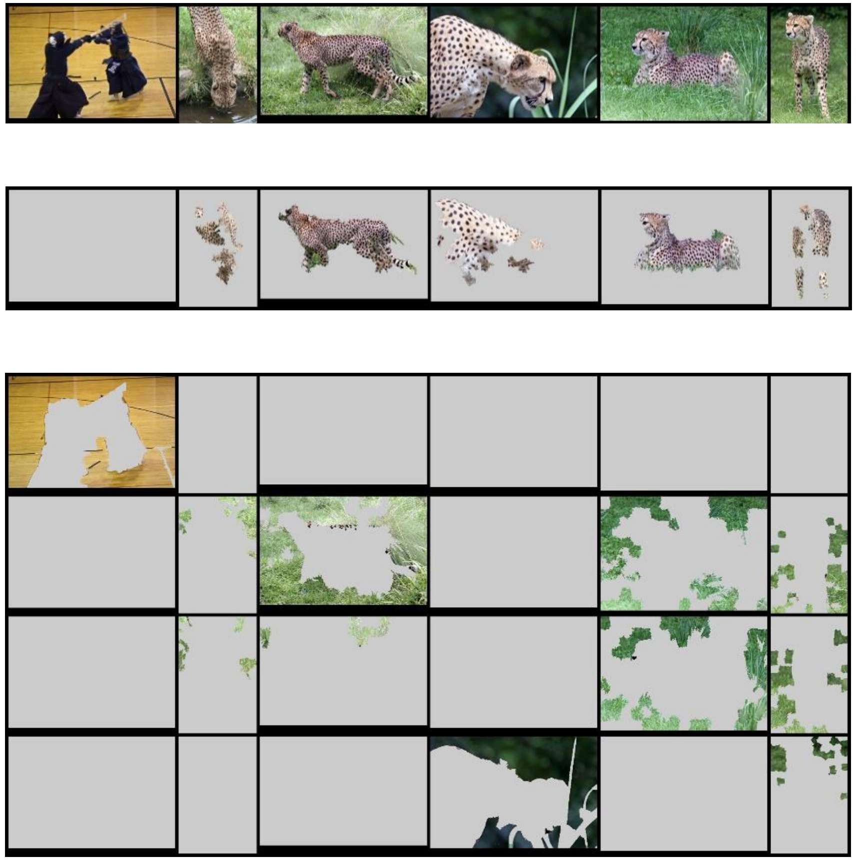
- Label assignment

$$\text{label}(\mathbf{s}_i) = \arg\max_j \mathbf{L}^*_{ij}$$

with spatial constraint



Proposed method

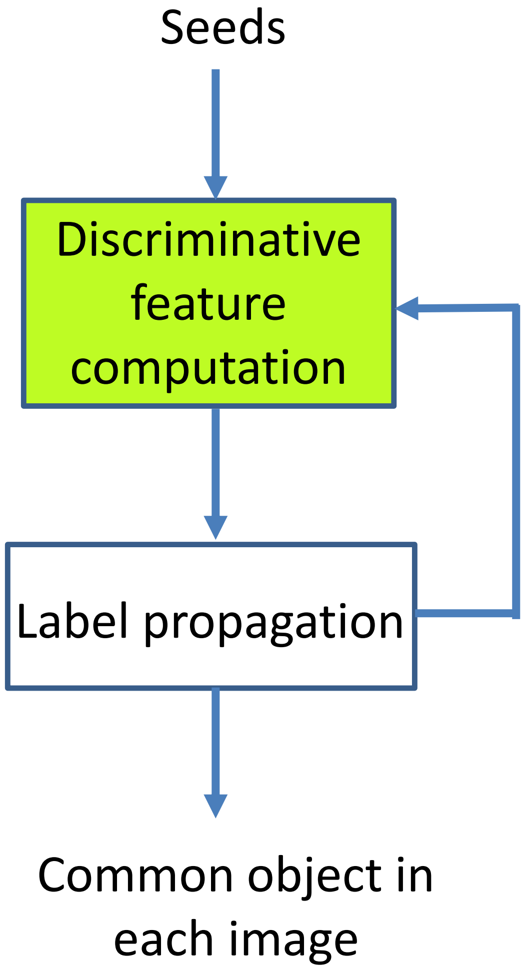


Input images

Common foreground seed using mode detection

Background seed

Proposed method



Seeds

Discriminative feature computation

Label propagation

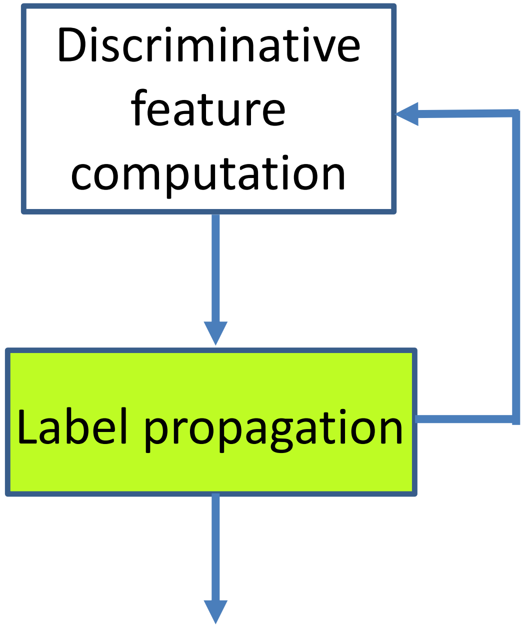
Common object in each image

Input images

Common foreground seed using mode detection

Background seed

Proposed method



Discriminative feature computation

Label propagation

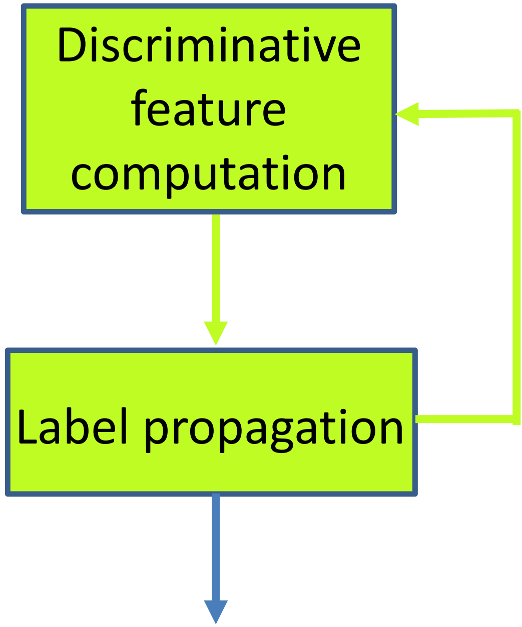
Common object in each image

Input images

Common foreground seed using mode detection

Background seed

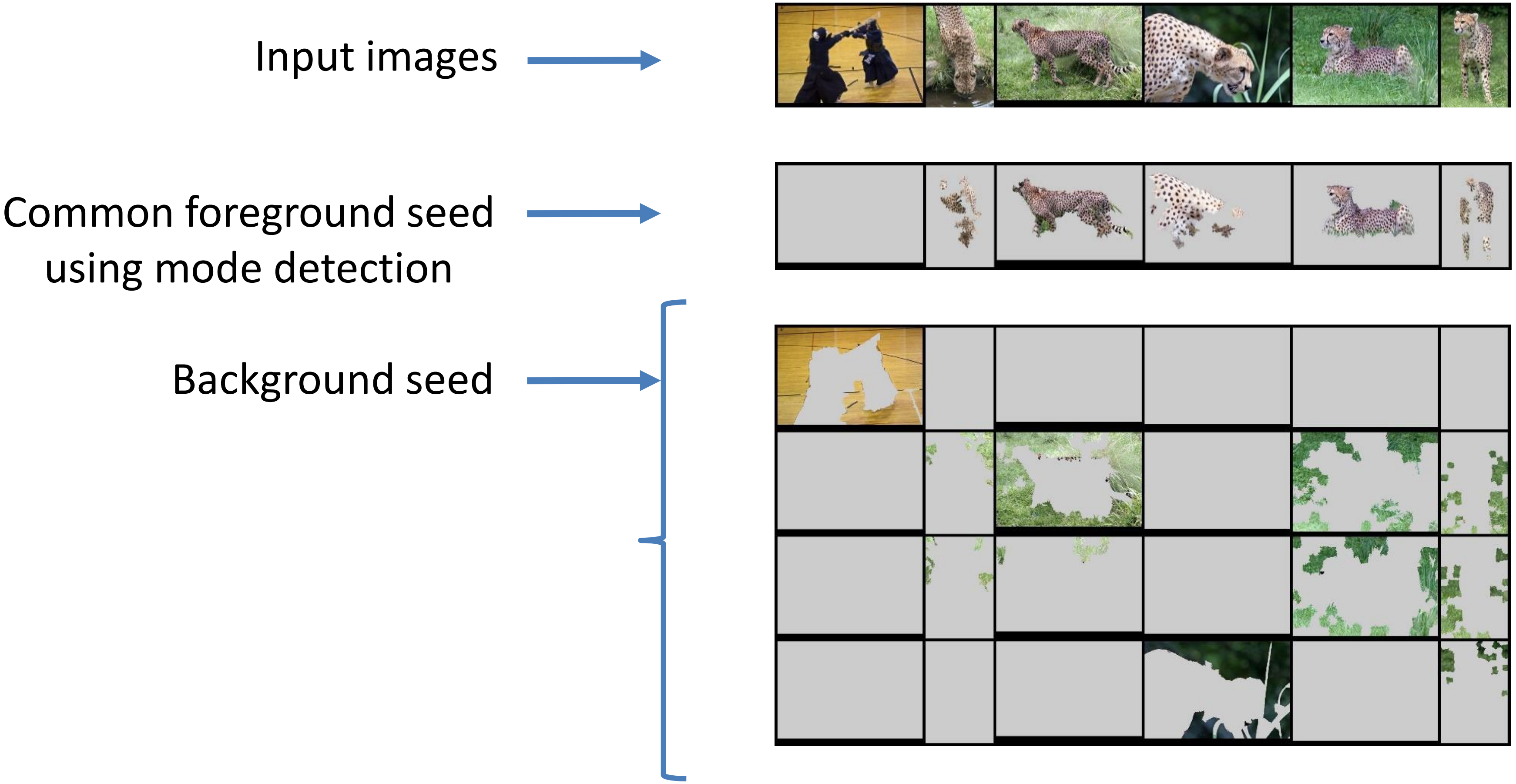
Proposed method



Discriminative feature computation

Label propagation

Common object in each image

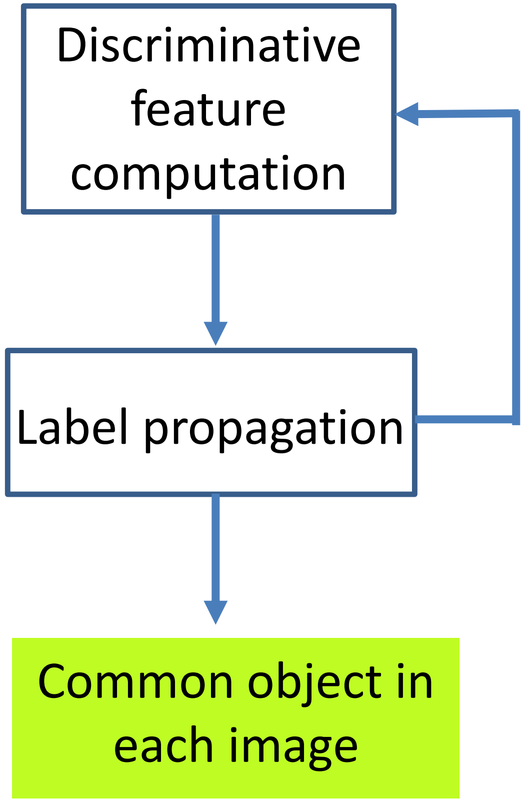


Input images

Common foreground seed using mode detection

Background seed

Proposed method



Discriminative feature computation

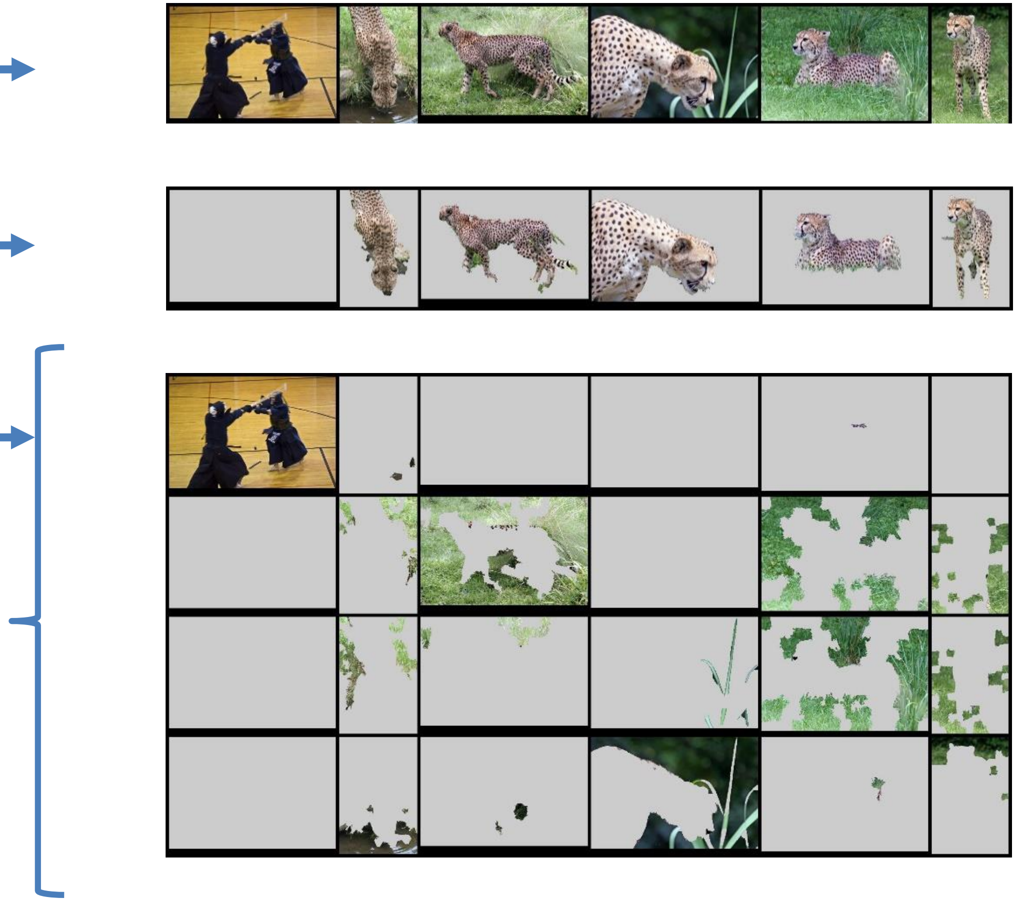
Label propagation

Common object in each image

Input images

Updated common foreground

Updated background



Visual comparison

Input images

Rubinstein *et al.*, CVPR 2013

Joulin *et al.*, CVPR 2012

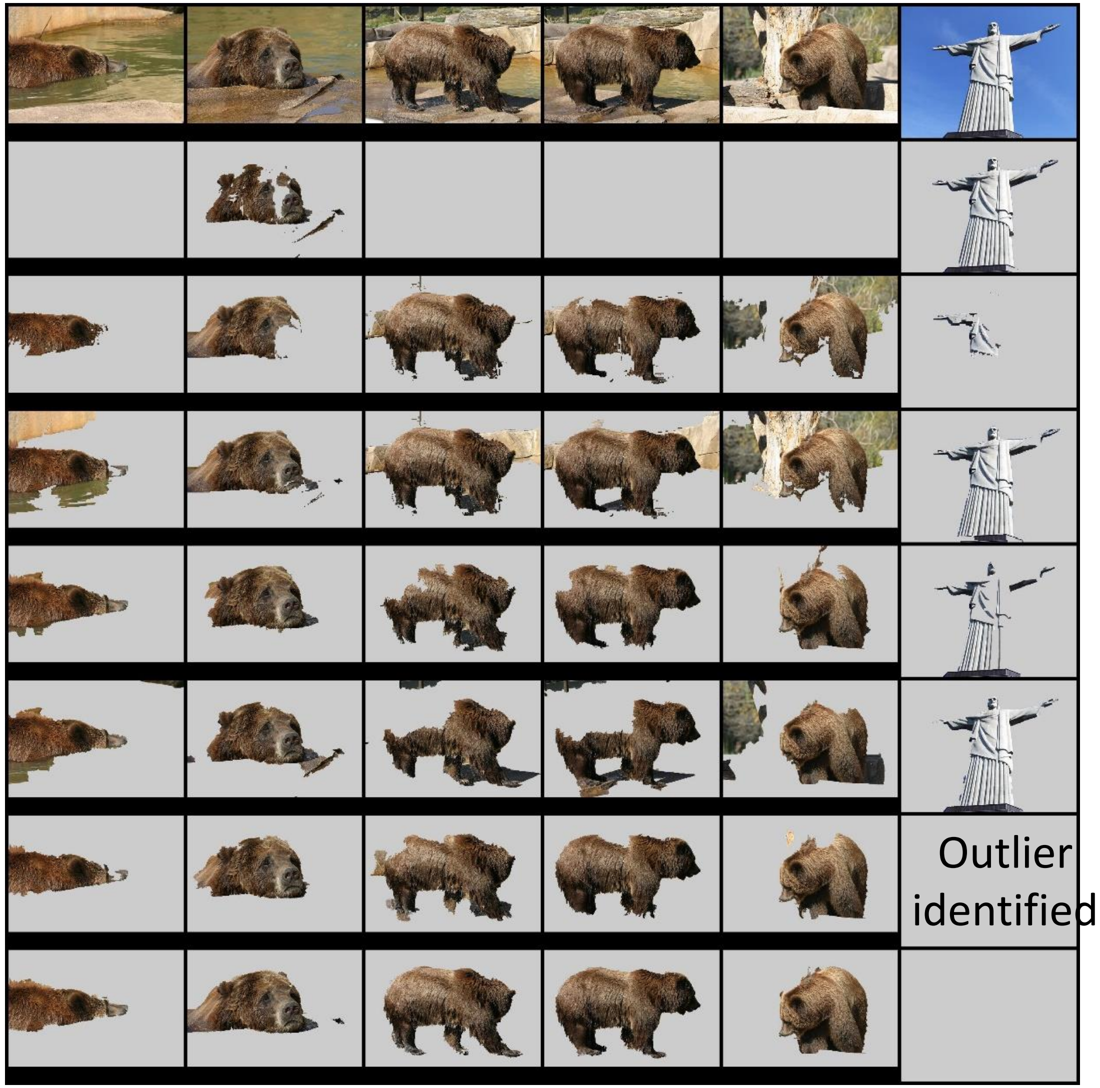
Joulin *et al.*, CVPR 2010

Lee *et al.*, CVPR 2015

Chang *et al.*, CVIU 2015

Proposed method

Ground-truth



Outlier identified

Image courtesy: iCoseg dataset

Visual comparison

Input images

Rubinstein *et al.*, CVPR 2013

Joulin *et al.*, CVPR 2012

Joulin *et al.*, CVPR 2010

Lee *et al.*, CVPR 2015

Chang *et al.*, CVIU 2015

Proposed method

Ground-truth

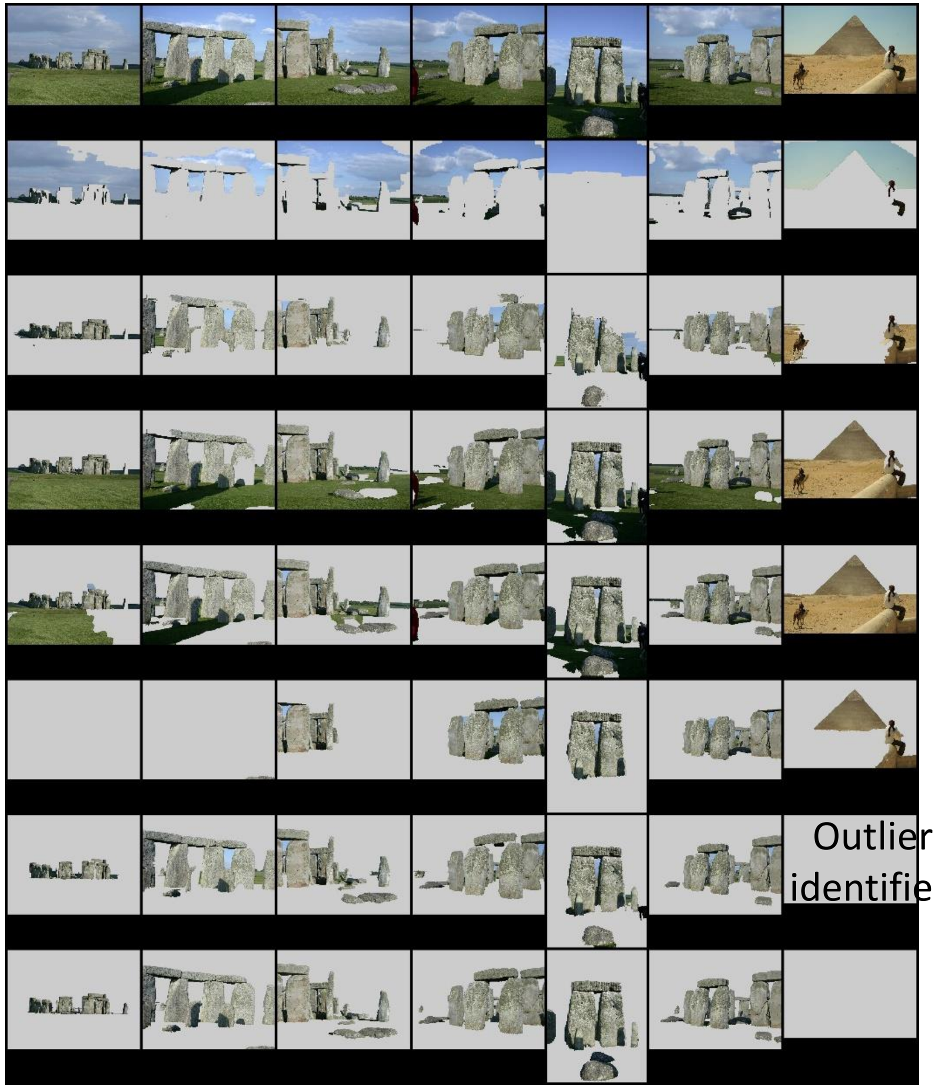
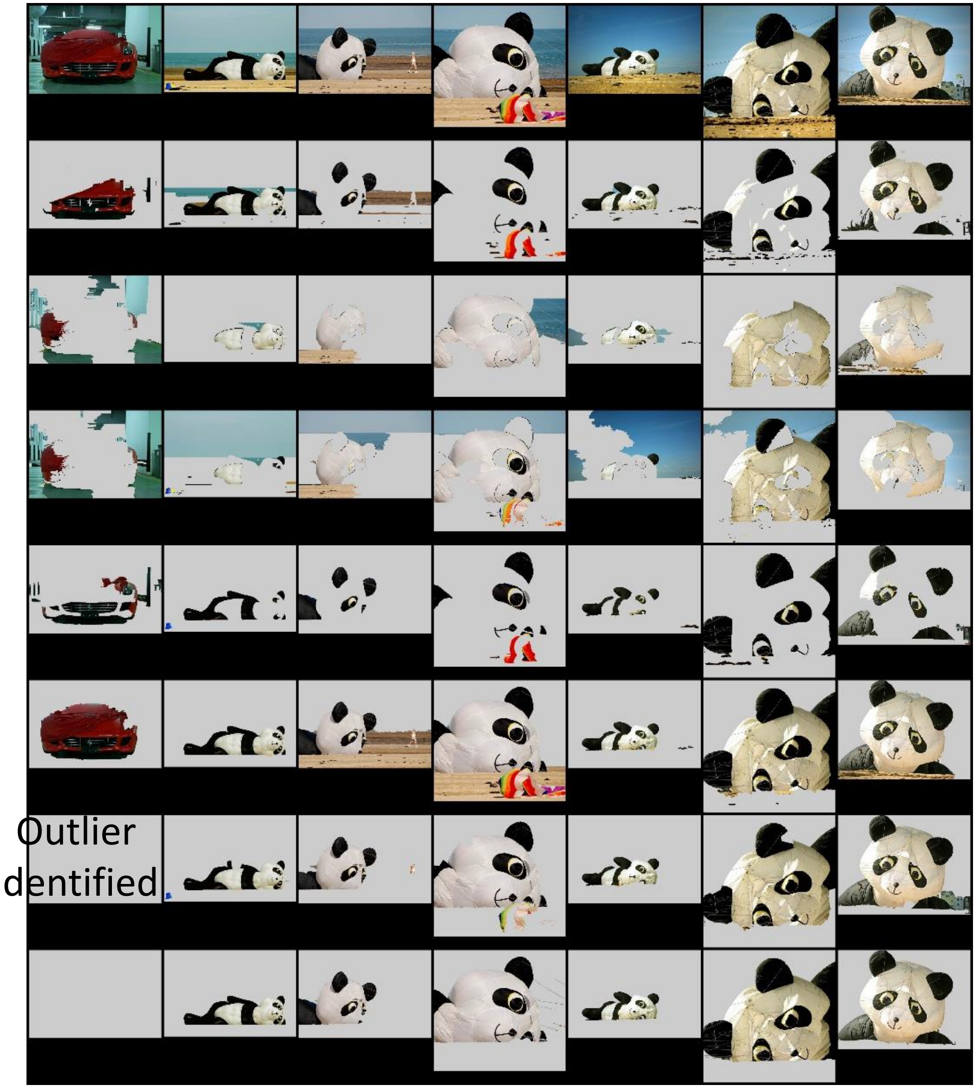


Image courtesy:
iCoseg dataset

Visual comparison



Input images

Rubinstein *et al.*, CVPR 2013

Joulin *et al.*, CVPR 2012

Joulin *et al.*, CVPR 2010

Lee *et al.*, CVPR 2015

Chang *et al.*, CVIU 2015

Proposed method

Outlier identified

Ground-truth

Image courtesy:
iCoseg dataset

Quantitative results

- Datasets
  - 570 outlier image sets created using images from iCoseg dataset
  - MIT object discovery set of internet images

Methods / Jaccard	PM	A	B	C	D	E	F
iCoseg 570	**0.71**	0.64	0.62	0.39	0.37	0.33	0.23
Internet dataset	**0.54**	0.44	0.40	-	-	-	-

PM : Proposed method

A : Lee *et al.*, CVPR 2015

B : Chang *et al.*, CVIU 2015

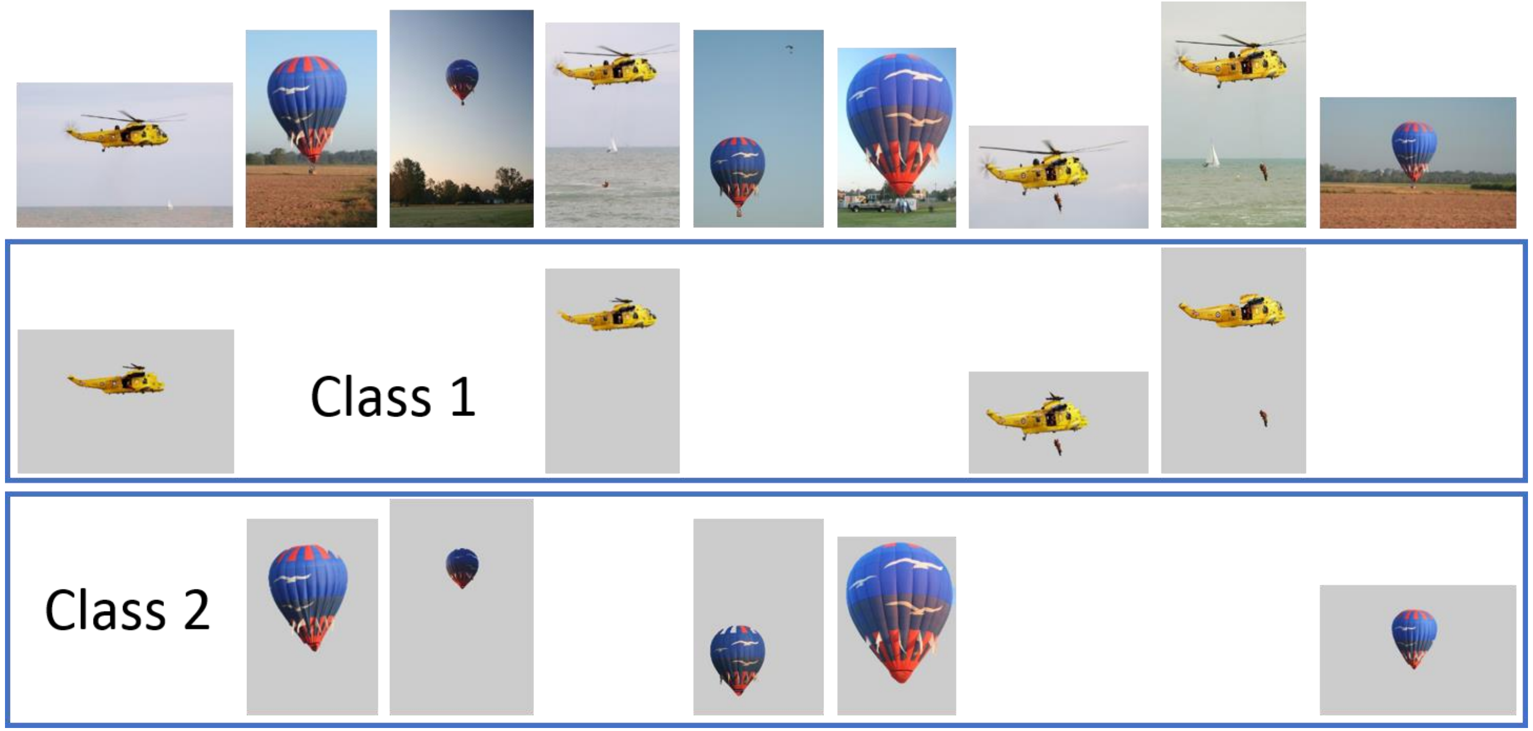
C : Joulin *et al.*, CVPR 2012

D : Rubinstein *et al.*, CVPR 2013

E : Joulin *et al.*, CVPR 2010

F : Kim *et al.*, ICCV 2011

Multiple class co-segmentation



Most compact cluster

$$p_1 = \arg\max_{j} \left\{ \Gamma_j : \forall j \right\}$$

Second most compact cluster

$$p_2 = \arg\max_{j} \left\{ \Gamma_j : \forall j \setminus p_1 \right\}$$

Row 1: input set of 9 images containing objects of two classes
Row 2: common object of class 1 (Helicopter)
Row 3: common object of class 2 (Balloon)

Image courtesy:
iCoseg dataset

Conclusion

- Unsupervised method : feature selection is important
- More number of background classes helps
- Discriminative features using LDA
  - better separation between classes
- Label updating
  - corrects seeds selected without supervision

Thank You