Unsupervised Image Matching and Object Discovery as Optimization

Huy V. Vo[1,2,3], Françis Bach[1,2], Minsu Cho[4], Kai Han[5], Yann LeCun[6], Patrick Pérez[3] and Jean Ponce[1,2]

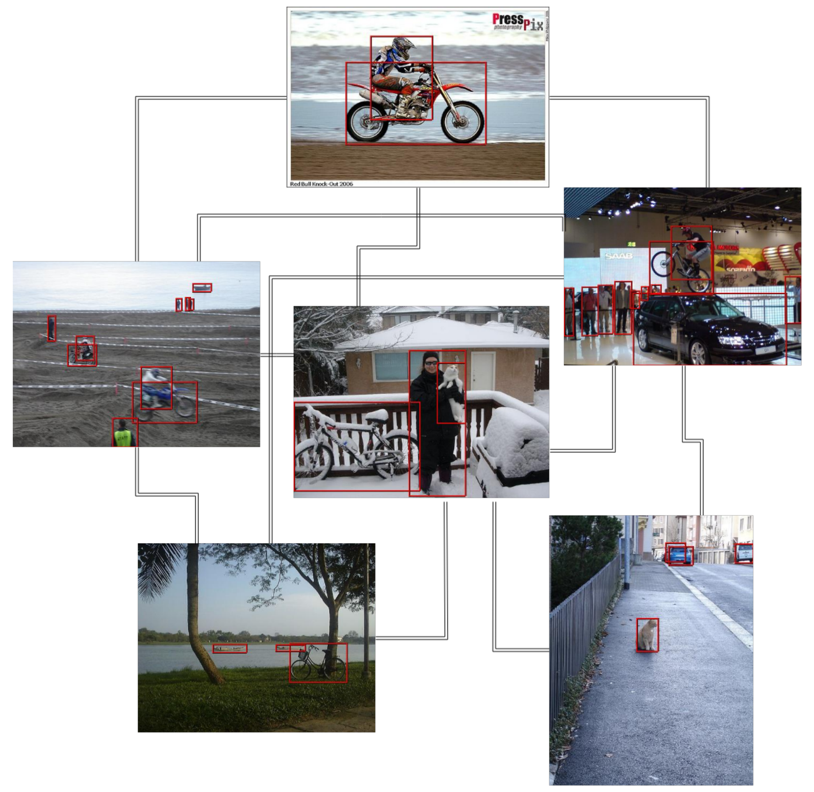
[1]Département d'informatique de l'ENS, ENS, CNRS, PSL University, Paris, France [2]INRIA, Paris, France [3]Valeo.ai [4]POSTECH

[5]University of Oxford [6]New York university

LONG BEACH CALIFORNIA June 16-20, 2019

OBJECT DISCOVERY

- Goal: Finding locations of foreground objects in images without any form of supervision.
- Motivation:
  - A fundamental problem in Computer Vision.
  - Cho *et al.* [1] propose a promising method but it is just a heuristic.
- The graph of images:



CONTRIBUTION

- A new saliency score for region matches between images.
- Reformulating Unsupervised Object Discovery as an optimization problem.
- Empirically proving that the new method gives better results.

REFERENCES

1 M. Cho, S. Kwak, C. Schmid and J. Ponce. Unsupervised object discovery and localization in the wild: Part-based matching with bottom-up region proposals. In CVPR, 2015.

2 Y. Li, L. Liu, C. Shen and A. Hengel. Image co-localization by mimicking a good detector's confidence score distribution. In ECCV, 2016.

3 X. Wei, C. Zhang, Y. Li, C. Xie, J. Wu, C. Shen and Z. Zhou. Deep descriptor transforming for image co-localization. In IJCAI, 2017.

4 S. Manen, M. Guillaumin and L. Van Gool. Prime object proposals with randomized Prim's algorithm. In ICCV, 2013.

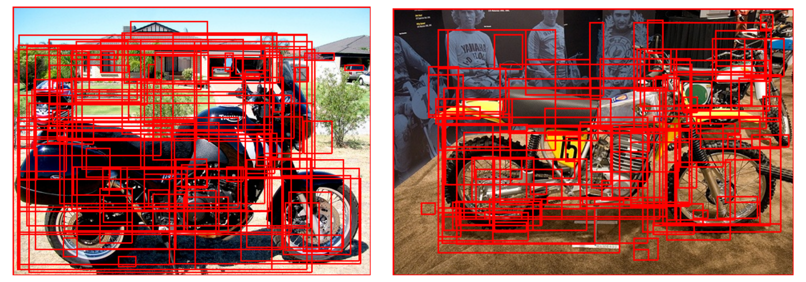
5 J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers and A. W. M. Smeulders. Selective search for object recognition. In IJCV, 2013.

6 B. Hariharan, J. Malik and D. Ramanan. Discriminative decorrelation for clustering and classification. In ECCV, 2012.

7 K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.

OBJECT DISCOVERY AS OPTIMIZATION

- Region proposals:



- Objective function:
  - Image i is represented by a binary vector x_i of length p_i.
  - The neighborhood of image i is represented by a binary vector e_i of length n.
  - There is a matrix S_{ij} of size $p_i \times p_j$ containing the saliency of matches between images i and j.
  - Objective function:

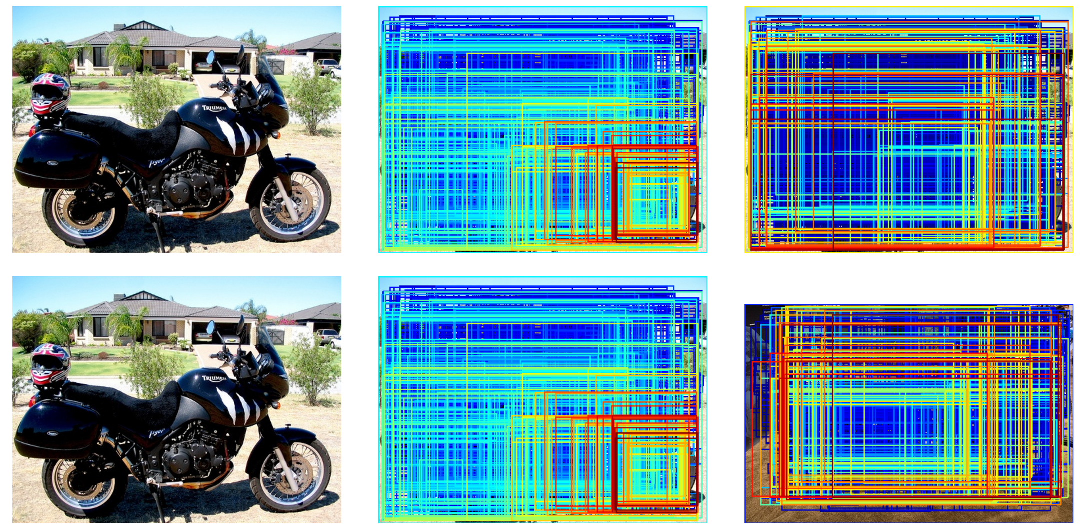
$$S(x,e) = \sum_{i \neq j} e_{ij} \sum_{\substack{1\le k\le p_i \\ 1\le l\le p_j}} S_{ij}^{kl} x_i^k x_j^l = \sum_{i\neq j} e_{ij}\langle x_i, S_{ij} x_j\rangle.$$

  - Constrains: $x_i \cdot \mathbf{1}_{p_i} \le \nu$ and $e_i \cdot \mathbf{1}_n \le \tau$.
- Similarity model:
  - Confidence score [1] measures the appearance similarity and geometric compatibility between regions

$$s_{ij}^{kl} = \frac{a_{ij}^{kl}}{p_i p_j} \sum_{\substack{1\le k'\le p_i \\ 1\le l'\le p_j}} K_{ij}^{kl,k'l'} a_{ij}^{k'l'}.$$

  - Based on the confidence score, standout score measures saliency of matches between pairs of regions

$$S_{ij}^{kl} = s_{ij}^{kl} - \max_{B_i^k \times B_j^l} s_{ij}^{kl}.$$



SOLVING THE OPTIMIZATION PROBLEM

- Objective as a concave function:

$$S(x,e) = \sum_{i\neq j} \sum_{\substack{1\le k\le p_i \\ 1\le l\le p_j}} S_{ij}^{kl} \min(e_{ij}, x_i^k, x_j^l).$$

- Lagrangian:

$$K(x,e;\lambda,\mu) = S(x,e) - \sum_{i}^{n} [\lambda_i(x_i \cdot \mathbf{1}_{p_i} - \nu) + \mu_i(e_i \cdot \mathbf{1}_n - \tau)].$$

- Dual problem: $\inf_{\lambda,\mu} \sup_{(x,e)\in[0,1]^*} K(x,e;\lambda,\mu)$.
- Solving the dual problem:

$$\begin{cases} \lambda_i^{(t+1)} = [\lambda_i^{(t)} + \alpha(x_i^{(t)} \cdot \mathbf{1}_{p_i} - \nu)]_+, \\ \mu_i^{(t+1)} = [\mu_i^{(t)} + \beta(e_i^{(t)} \cdot \mathbf{1}_n - \tau)]_+, \end{cases}$$

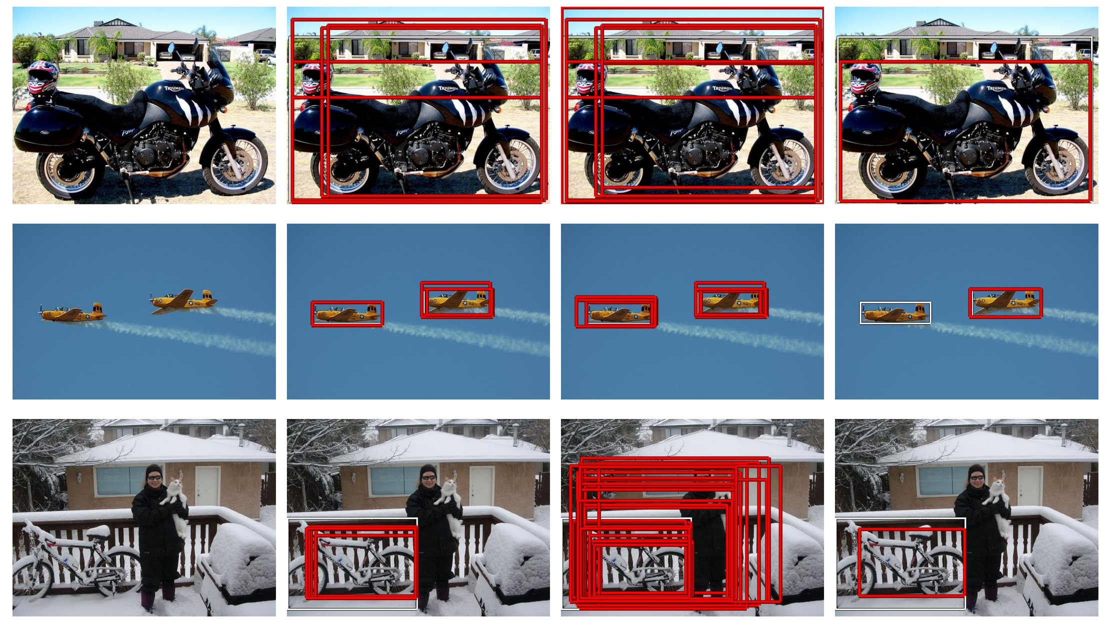
where

$$(x^{(t)}, e^{(t)}) = \operatorname*{argmax}_{(x,e)\in\{0,1\}^*} K(x,e;\lambda^{(t)},\mu^{(t)}).$$

- Approximate primal solution:

$$(x,e) = \frac{1}{N}\sum_{t=0}^{N-1} (x^{(t)}, e^{(t)}).$$

- Rounding and greedy coordinate ascent algorithm:
  - In a random order of x_i, update x_i to maximize $S(x,e)$ keeping other variables fixed.
  - Update e_i in parallel to maximize $S(x,e)$ keeping x fixed.
- Post processing ensembling method.



RESULTS

- Datasets: Object Discovery, VOC 6x2, VOC all.
- Metric: *correct localization* (CorLoc).
- Results in the separate setting:

Method	OD	VOC 6x2	VOC all
Cho *et al.* [1]	84.2	67.7	36.6
Cho *et al.* (Our execution)	84.2	67.6	37.6
Li *et al.* [2]	-	-	40.0
Wei *et al.* [3]	**88.1**	-	**46.9**
Ours	85.8 ± 0.6	69.4 ± 0.3	39.2 ± 0.2

- Results in the mixed setting:

Method	OD	VOC 6x2	VOC all
Cho *et al.* [1]	-	-	37.6
Cho *et al.* (Our execution)	82.2	55.9	37.5
Ours	**83.0 ± 0.4**	**60.2 ± 0.4**	**39.8 ± 0.2**

- Results with proposals from selective search:

Proposals	selective search [4]			randomized Prim's [5]
Cho *et al.* [1]	23.3	20.6	32.6	67.6
Ours	**41.4 ± 0.5**	**48.4 ± 0.5**	**62.8 ± 0.6**	**69.4 ± 0.4**

- Results with CNN features:

Features	OD	VOC 6x2	VOC all
WHO [6]	**85.8 ± 0.6**	69.4 ± 0.3	39.2 ± 0.2
CNN [7]	78.8 ± 0.4	**70.9 ± 0.2**	**42.5 ± 0.1**

FUTURE WORK

- Symmetric version.
- Multiple objects.

Acknowledgments: This work was supported in part by the Inria/NYU collaboration agreement, the Louis Vuitton/ENS chair on artificial intellgence and the EPSRC Programme Grant Seebibyte EP/M013774/1. We also thank Simon Lacoste-Julien for his valuable comments and suggestions.