Large-Scale Unsupervised Object Discovery

Huy V. Vo[1,2], Elena Sizikova[3], Cordelia Schmid[1], Patrick Pérez[2] and Jean Ponce[1,3]

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

NEURAL INFORMATION PROCESSING SYSTEMS

INTRODUCTION

- **Goal**: Identify objects in a large collection of unlabelled images.
- **Motivation**:
  - Automatically analyze image collections.
  - Obtain free image annotations.
- **Recent Prior Work**:
  - Find objects that are frequently appearing visual patterns [8, 66, 67] using a sequential algorithm:

$$\max_{x,e} \sum_{p=1}^{n} \sum_{q \in \mathcal{N}(p)} e_{pq} x_p^T S_{pq} x_q$$

$$\text{s.t.} \sum_{k=1}^{r} x_p^k \le \nu \text{ and } \sum_{q \neq p} e_{pq} \le \tau \quad \forall p.$$

  - Limitations: Does not scale well, relies on supervised features for good performance.

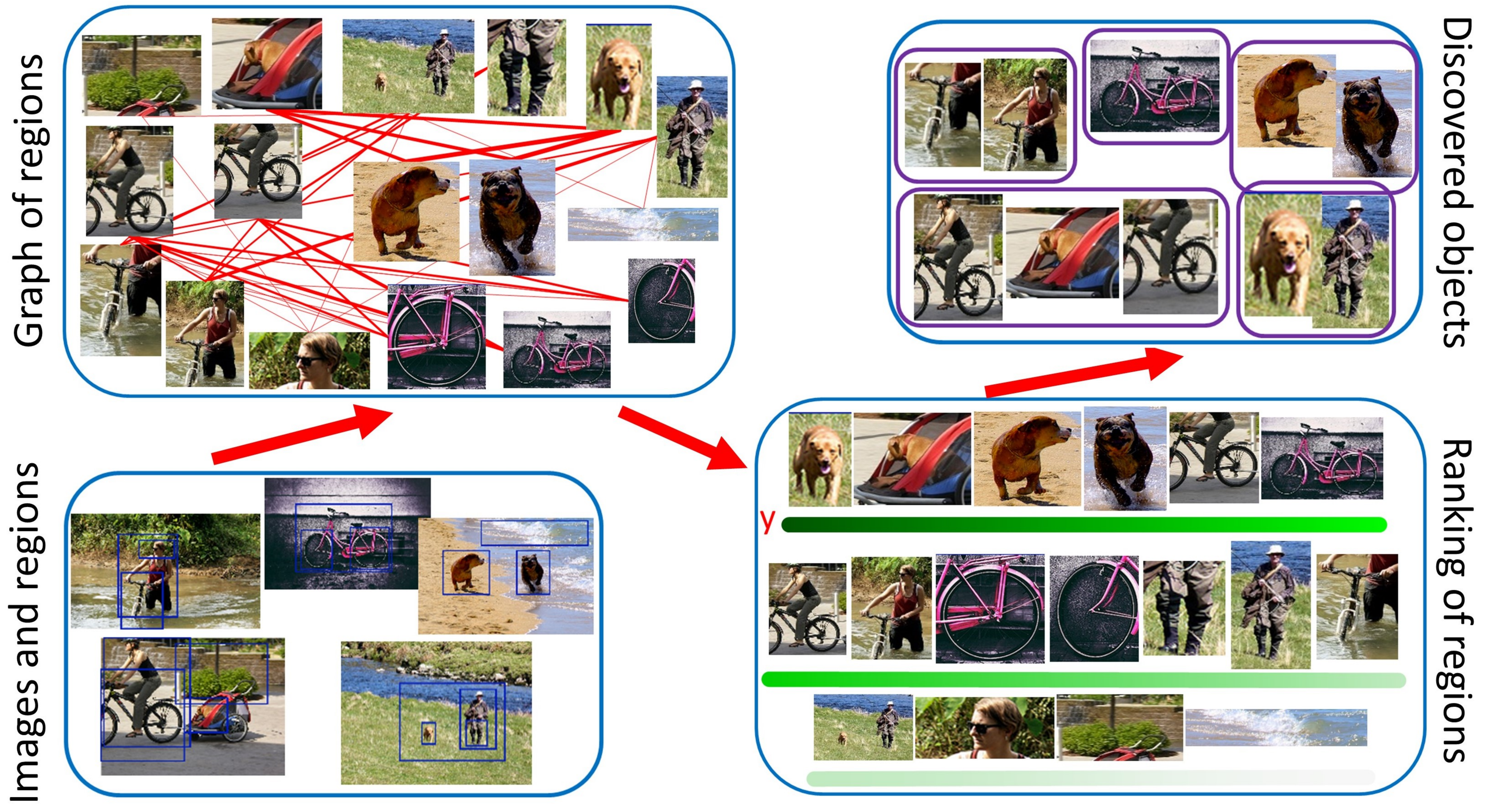
CONTRIBUTIONS

- We formulate UOD as a ranking problem to:
  - allow parallel and distributed solutions.
  - scale UOD up to datasets 87 times larger than those considered in the previous work.
  - improve multi-object discovery performance by up to 32% according to Average Precision.
- We demonstrate the first viable pipeline for completely unsupervised object discovery using self-supervised features.

References: [8] Cho et al., CVPR'15; [14] Feng et al., ICCV'11; [31] Kim et al., CVPR'08; [32] Kim et al., NIPS'09; [33] Kim et al., CVPR'12; [35] Krasin et al., 2017; [37] Langville et al., Internet Mathematics, 2004; [42] Lin et al., ECCV'14; [48] Page et al., 1999; [66] Vo et al., CVPR'19; [67] Vo et al., ECCV'20; [68] Mises et al., 1929; [71] Wei et al., PR'19; [73] Zhang et al., Trans. On Neural Networks, 2011; [74] Zhang et al., Applied Intelligence, 2009; [75] Zhang et al., ICCV'15; [82] Zhu et al., CVPR'12; [83] Zitnick et al., ECCV'14.

Acknowledgments: This work was supported in part by the Inria/NYU collaboration, the Louis Vuitton/ENS chair on artificial intelligence and the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR19-P3IA-0001 (PRAIRIE 3IA Institute). Elena Sizikova was supported by the Moore-Sloan Data Science Environment initiative (funded by the Alfred P. Sloan Foundation and the Gordon and Betty Moore Foundation) through the NYU Center for Data Science. Huy V. Vo was supported in part by a Valeo/Prairie CIFRE PhD Fellowship. We thank Spyros Gidaris for providing the VGG16-based OBoW model. Finally, we thank anonymous reviewers for their helpful suggestions and feedback for the paper.

OVERVIEW



PROPOSED RANKING FORMULATIONS

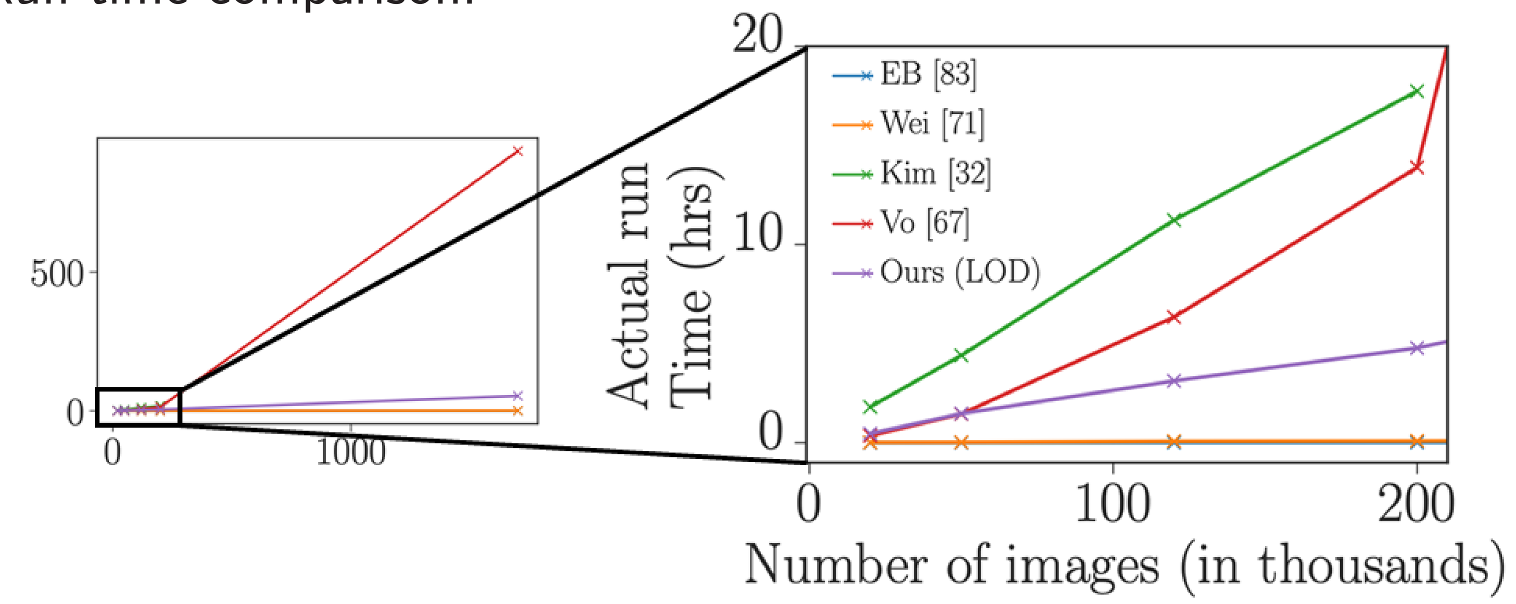
- **Input**: Symmetric weight matrix W, # of nodes N.
- **Output**: Nodes' importance.
- A node is important if it is well-connected and its neighbors are important.
- **Quadratic optimization (Q)**:
  - Importance score: $y_i \in [0,1], y = (y_1, y_2, \ldots, y_N)^t$.
  - Total importance of edges of node i: $c_i = \sum_j y_i W_{ij} y_j$.
  - Optimization: $y^* = \underset{y \ge 0, \|y\| \le 1}{\text{argmax}} \sum_i c_i = y^t W y$.
  - y^* is the largest eigenvector of W.
- **PageRank (P)** [37,48]:
  - Nodes are states of a Markov chain. Transition matrix: $A = W\text{diag}(\mathbb{1}_N^t W)^{-1}$.
  - PageRank matrix: $P = (1-\beta)A + \beta u \mathbb{1}_N^t$ with $\beta \in [0,1]$ and $u \ge 0$ is the *personalized vector*.
  - PageRank vector: The largest eigenvector of P.
- **Hybrid formulation (LOD)**:
  - Q's solution is good. Using Q to build a personalized vector $u(Q)$.
  - LOD: Personalized PageRank with $P = (1-\beta)A + \beta u(Q)\mathbb{1}_N^t$.
- (Q), (P) and (LOD) are all eigenvector problems solved with a distributed Power iteration implementation [68].

EXPERIMENTAL RESULTS

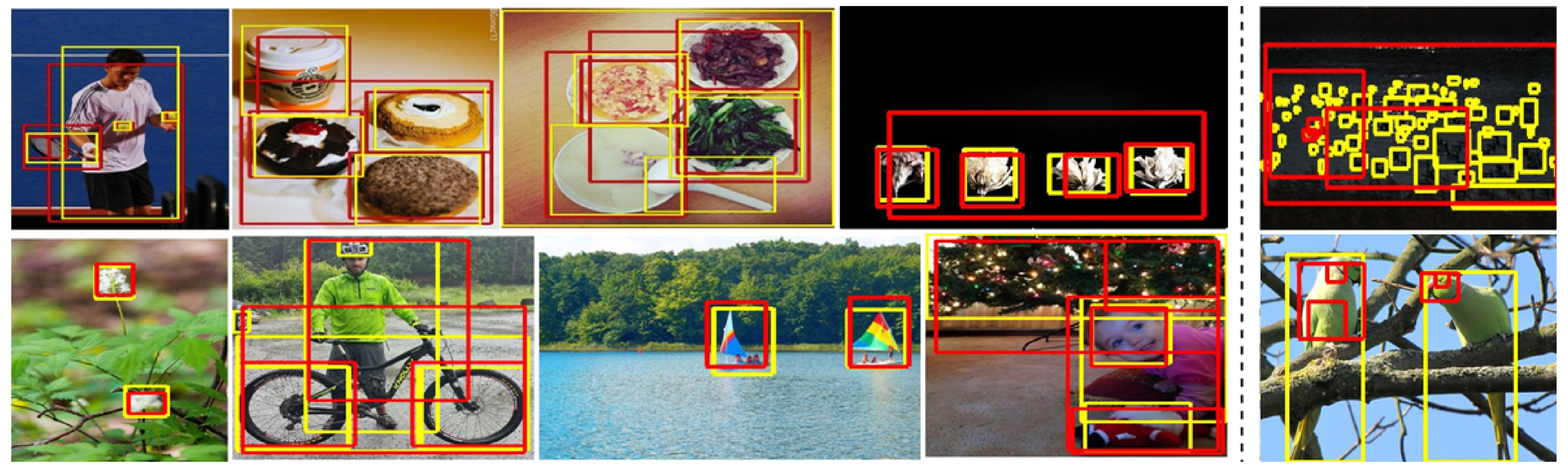
- **Datasets**: COCO trainval [42] (120k images), COCO20K (20k images), OpenImages [35] (1.7 million images), Open50K (50k images).
- **Evaluation Metrics**: correct localization (CorLoc, precision of returned boxes in single-object setting) and average precision (AP).
- Quantitative comparison to the state of the art:

Method	Single-object				Multi-object							
	CorLoc (↑)				AP50 (↑)				AP@[50:95] (↑)			
	C20K	C120K	Op50K	Op1.7M	C20K	C120K	Op50K	Op1.7M	C20K	C120K	Op50K	Op1.7M
EB [83]	28.8	29.1	32.7	32.8	4.86	4.91	5.46	5.49	1.41	1.43	1.53	1.53
Wei [71]	38.2	38.3	34.8	34.8	2.41	2.44	1.86	1.86	0.73	0.74	0.6	0.6
Kim [32]	35.1	34.8	37.0	-	3.93	3.93	4.13	-	0.96	0.96	0.98	-
Vo [67]	**48.5**	48.5	48.0	47.8	5.18	5.03	4.98	4.88	1.62	1.6	1.58	1.57
Ours (LOD+Self)	41.1	42.4	**49.5**	**49.4**	4.56	4.90	6.37	**6.28**	1.29	1.37	1.87	**1.86**
Ours (LOD)	**48.5**	**48.6**	48.1	47.7	**6.63**	**6.64**	**6.46**	**6.28**	**1.98**	**2.0**	**1.88**	1.83

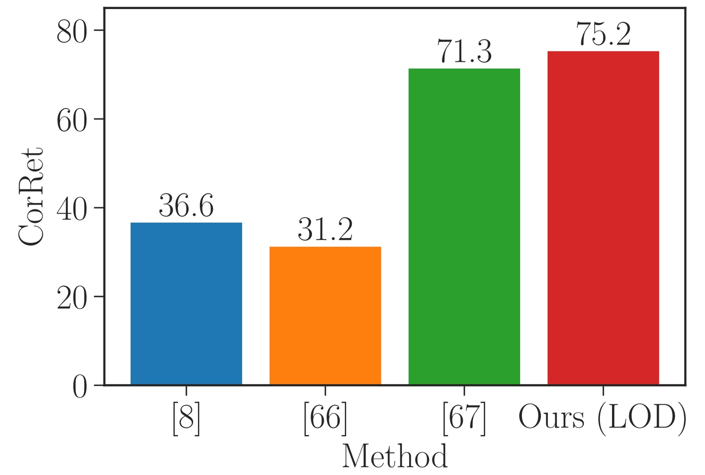
- Run time comparison:



- Sample qualitative results:



- Object category discovery:



Dataset	LOD	[75]	[82]	[14]	[73]	[74]	[31]	[33
SIVAL1	**97.4**	89.0	95.3	80.4	39.3	38.0	27.0	45.0
SIVAL2	**99.0**	93.2	84.0	71.7	40.0	33.3	35.3	33.3
SIVAL3	88.3	**88.4**	74.7	62.7	37.3	38.7	26.7	41.3
SIVAL4	**97.7**	87.8	94.0	86.0	33.0	37.7	27.3	53.0
SIVAL5	**94.3**	92.7	75.3	70.3	35.3	37.7	25.0	48.3
Average	**95.3**	90.2	84.7	74.2	37.0	37.1	28.3	44.2