Co-localization with Category-Consistent Features and Geodesic Distance Propagation

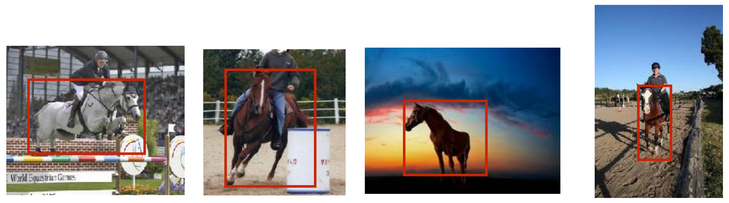
Hieu Le[1], Chen-Ping Yu[3], Gregory Zelinsky[1,2] & Dimitris Samaras[1]
[1]Department of Computer Science; [2]Department of Psychology; Stony Brook University.
[3] Phiar Technologies, Inc.

Stony Brook University

ICCV 17

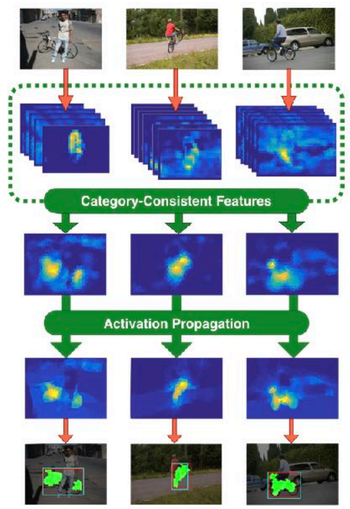
Introduction

Co-localization is the problem of localizing the common objects from a set of images.



We show that a pre-trained network can perform image co-localization without **fine-tuning** and without any **negative examples.**

Framework



Set of input images → Identify CCFs → Obtain objects' heat map activation → Activation propagation for better object localization

Contributions

- We propose a novel **feature extraction** method that can select a set of representative features for a category using only positive images.
- We propose an effective **activation propagation** method using superpixel geodesic distances to refine object segmentations from CNN heat maps.

CCFs selection

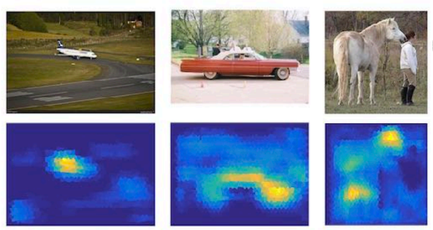
- Features are grouped using k-means (L-p distance) based on their similarity in activation values. Features (**A**,**B**) would be grouped, as would features (**C**,**D**). Feature (**E**) would remain ungrouped.
- CCFs are defined as the features (or feature singleton) in the group having the highest average activation.

We select a group of features that are highly activated across all images and have similar activations.

- CCFs visualization:

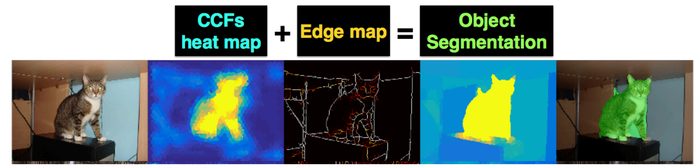
- Examples of CCFs selected for the "airplane" and "bus" category from VOC 2012. All CCFs are from the last convolutional layer of VGG19 [5].
- The first column shows the input exemplars, the other columns show the activated areas corresponding to the top 4 CCFs selected by our method.

- Heat maps from all CCFs combined:

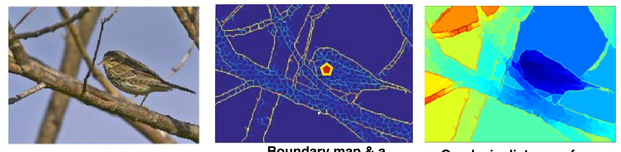


- Examples of combined activation from identified CCFs on three classes of Object Discovery dataset, namely "airplane", "car", & "horse".

Geodesic distance Propagation



- We employ geodesic distances, which compactly encode all information of the edge maps, for activation propagation.
- Given **a boundary probability map** [6] and **a superpixel segmentation** of an image, a graph ***G*** is defined as:
  - Each node is a superpixel.
  - Each edge connects adjacent superpixels.
  - The edge weight is the likelihood of an object boundary between two adjacent superpixels, computed from the boundary probability map
  - **Geodesic Distance between two superpixels:** The shortest path between two superpixels in ***G***



Input Image | Boundary map & a superpixel-of-interest | Geodesic distances from the superpixel-of-interest to other superpixels

- Each superpixel distributes its activation to all others superpixels. The propagation value between a superpixel ***i*** to superpixel ***j*** is defined by:

$$W_{i,j} = \frac{\exp(-d_{i,j} \times \mu^{-1})}{\sum_{k=1}^{N} \exp(-d_{i,k} \times \mu^{-1})}$$

- Simultaneously update superpixels' activations:

$$E' = \mathbf{WE}$$

Results

Dataset	Rubinstein et al. [4]	Joulin et al. [3]	Cho et al. [1]	Yi et al.[2]	Ours
VOC07	N.A	24.6	36.6	40.0	41.97
VOC12	N.A	N.A	41.8	43.8	48.22
Object Discovery	75.16	76.58	84.19	N.A	86.03
6 Imagenet subsets	N.A	N.A	37.72	48.12	60.95

Average CorLoc scores (0.5 IoU) of our approach and state-of-the-art methods on multiple datasets. Best scores are in red, second best are in blue.

ablation study: the effect of geodesic distance propagation

Dataset	Without GDP	With GDP
VOC07	35.41	41.97
VOC12	41.2	48.22

Two examples illustrating the effect of our feature selection method. The first row is the result of our method using the identified CCFs, the second row is the result using other features

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. van den Hengel. Image colocalization by mimicking a good detector's confidence score distribution. In ECCV, 2016.
[3] A. Joulin, K. Tang, and F.-F. Li. Efficient image and video co-localization with frank-wolfe algorithm. In ECCV, 2014.
[4] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu. Unsupervised joint object discovery and segmentation in internet images. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), 2013.
[5] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.
[6] Larry Zitnick Piotr Dollar. Structured forests for fast edge detection. In ICCV. International Conference on Computer Vision, December 2013.

Acknowledgement
This work was partially supported by the Vietnam Education Foundation, the Stony Brook University SensorCAT, a gift from Adobe, the Partner University Fund 4DVision project, and New York State ITSC.