

SILCO – See a Few Images, Localize the Common Object

Tao HU, Pascal Mettes, Jia-Hong Huang and Cees G. M. Snoek
University of Amsterdam
ICCV 2019

Presented by Dax Borde

UCF

Table of Contents

- Problem Intro
- Related Work
- Method
- Results
- Conclusion

UCF



Problem Intro

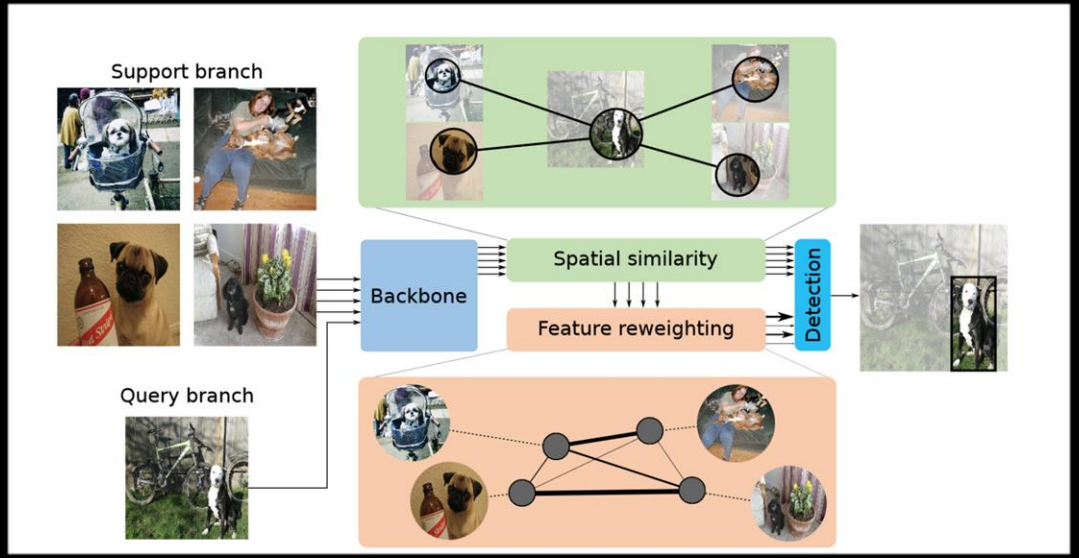
Problem Intro

Inputs

1. Several Support Images (3 - 9)
2. Query Image

Output

- Bounding Box in Query

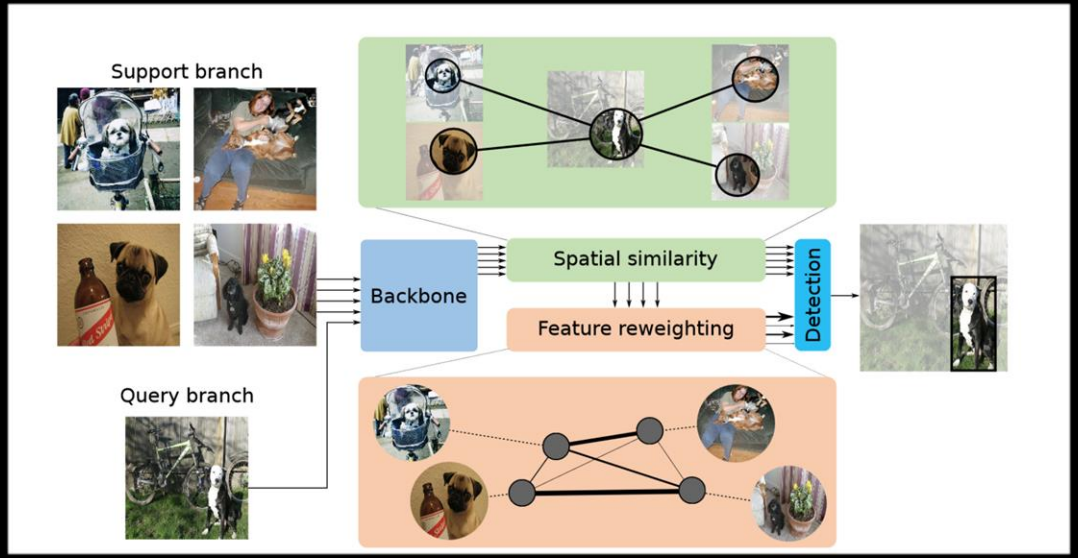


UCF

Problem Intro

Advantages

1. Less Data
2. Weak Supervision
3. Combination w/ others

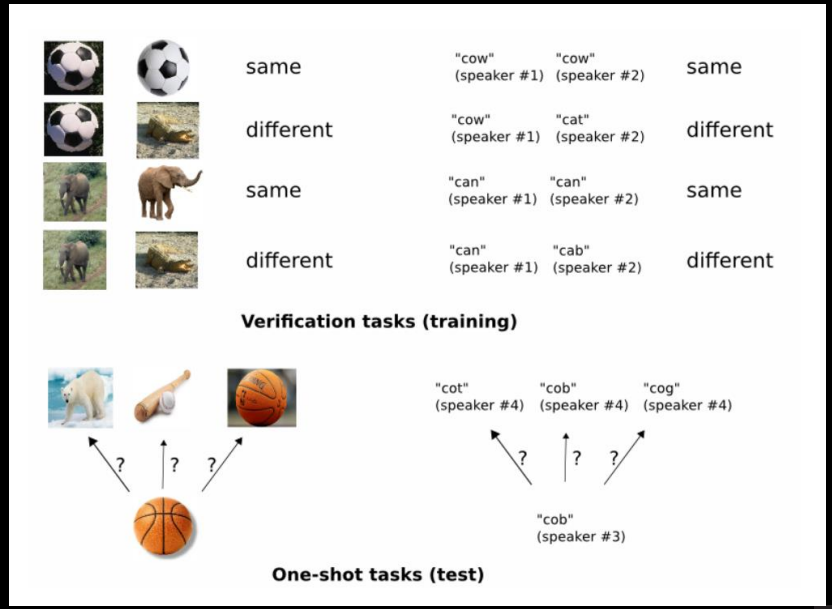


Related Work



Few Shot Classification

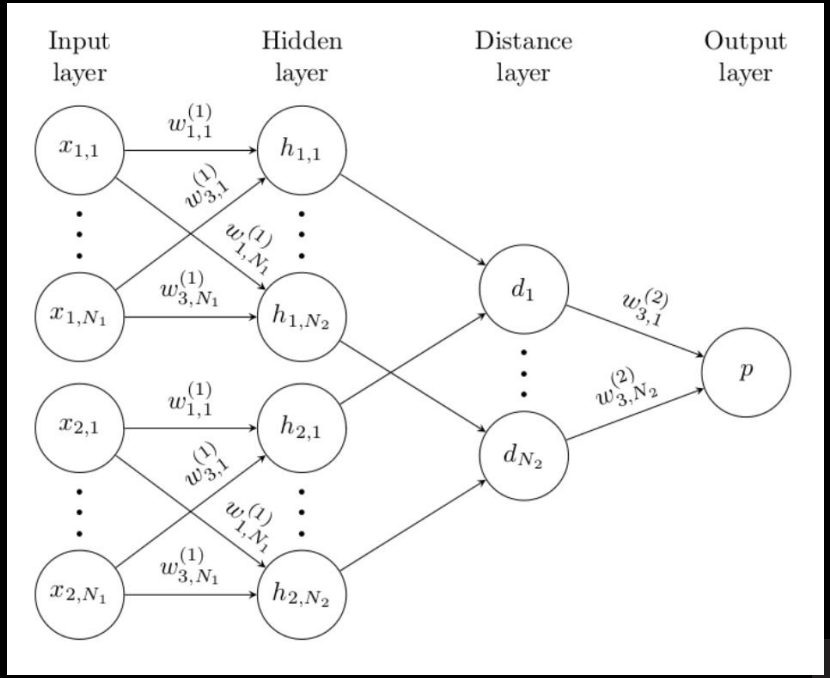
- Find similarities robustly
- Evaluate on novel classes
- Foundation for Similarity Mod



Few Shot Classification

Siamese Networks

- One shot only
- Fully connected
  - Inserted after convolutions
  - Weights tied together
- Learn similarity



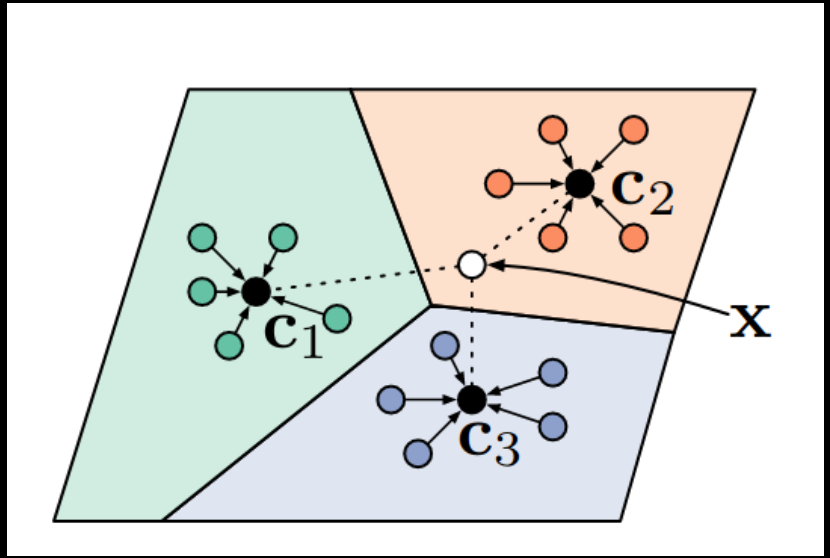
https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

UCF

Few Shot Classification

Prototypical Networks

- Few shot (or 0 shot)
- Predefined classes
- Embeddings



https://arxiv.org/pdf/1703.05175.pdf

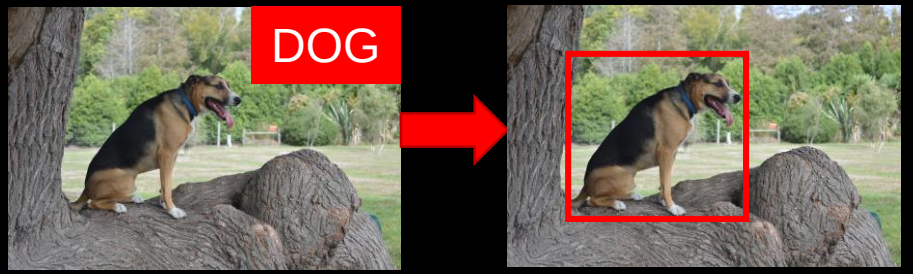
UCF

Graph Convolutional Networks

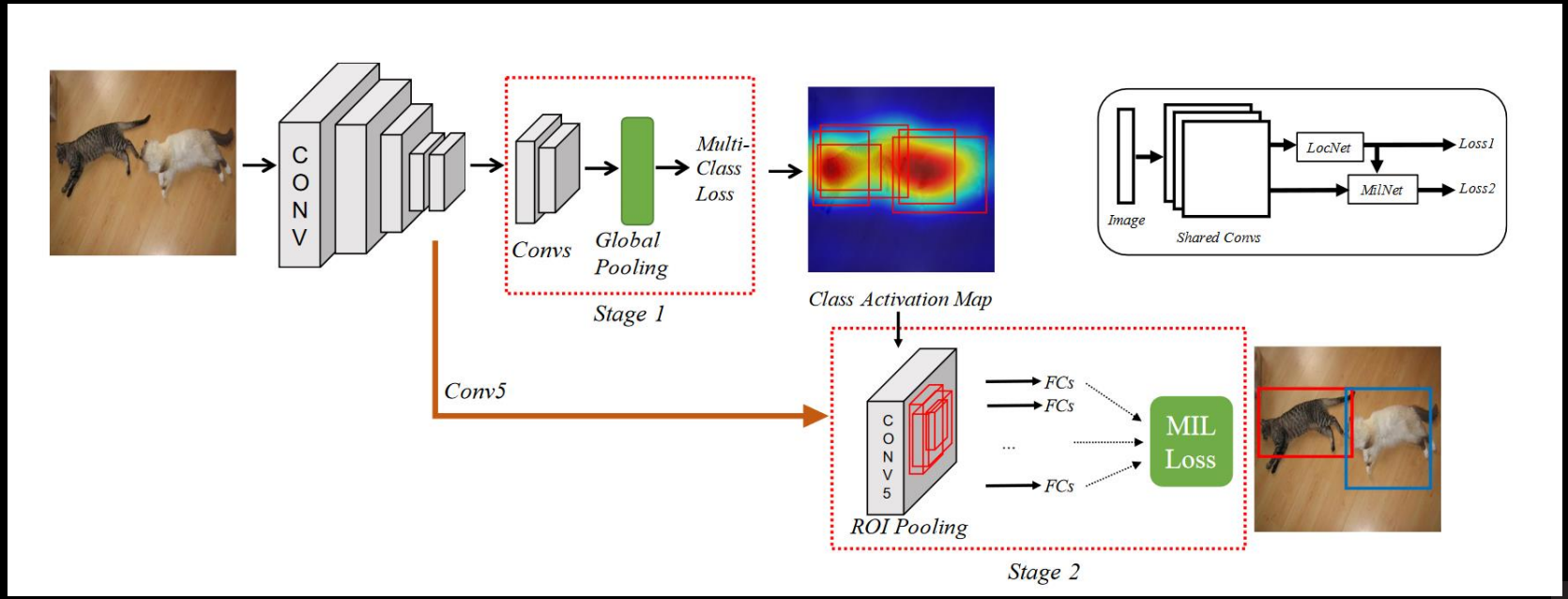
More information in Method section!

Weakly Supervised Detection

- Given image and label
- Localize labeled object
- Requires Lots of weak labels
- Predefined Classes



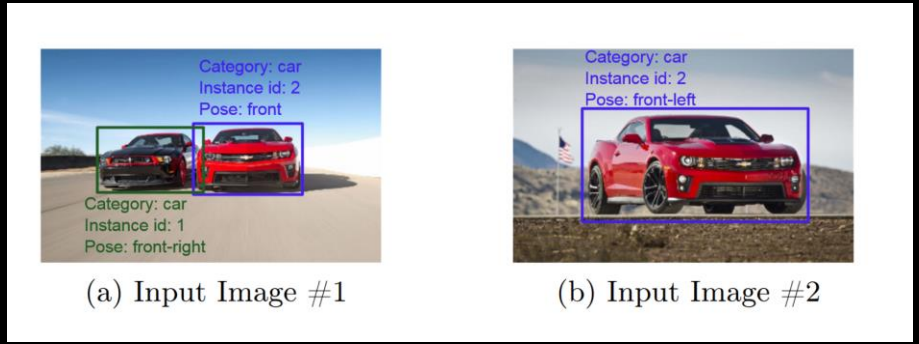
Weakly Supervised Detection



https://arxiv.org/pdf/1611.08258.pdf

Object Co-Detection

- Detect same object in 2 imgs
- Predefined Classes
- CRFs for sequential detection



https://cvgl.stanford.edu/papers/bao_eccv12_codetection.pdf http://users.cecs.anu.edu.au/~hexm/papers/codetect_eccv2014.pdf

UCF

How SILCO is Different

- Variable number of input images
- Less hunger for labeled data & bounding boxes
- No predefined classes

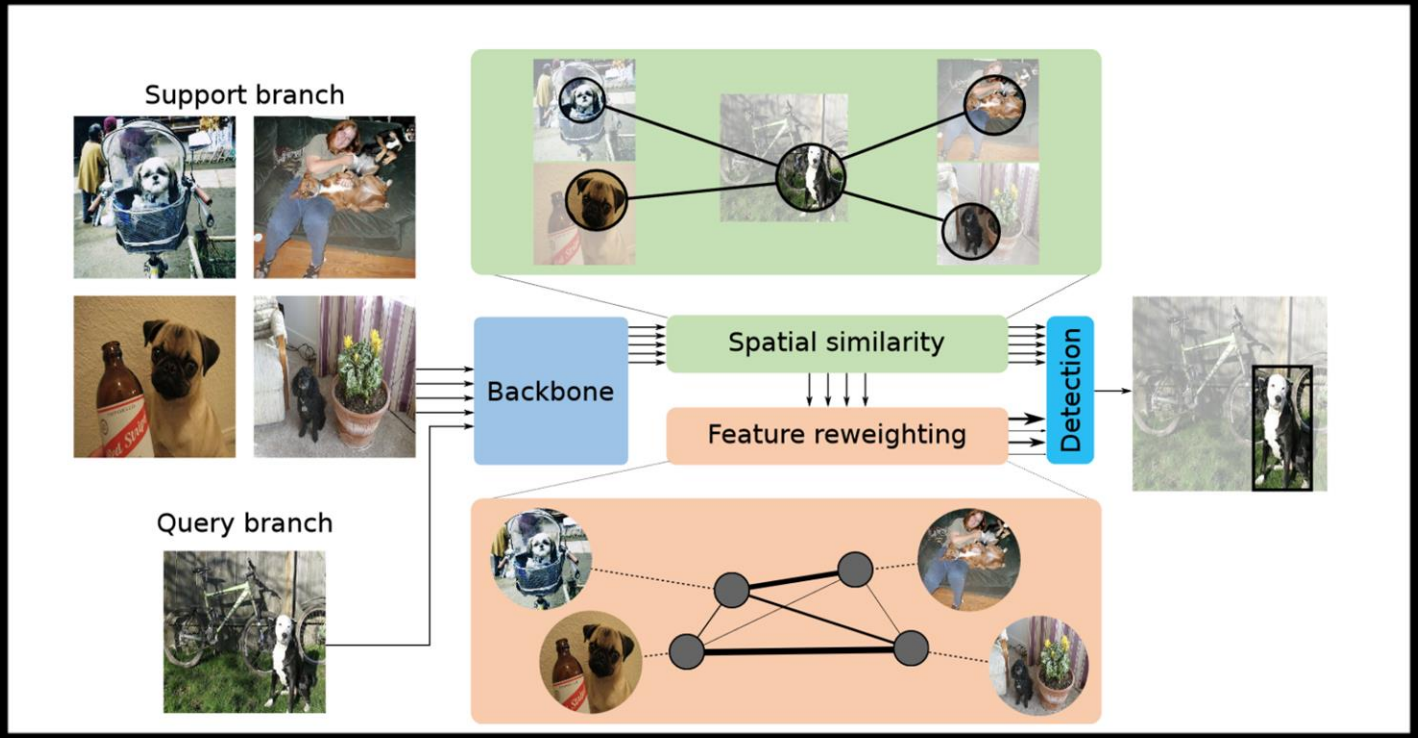


Method

UCF

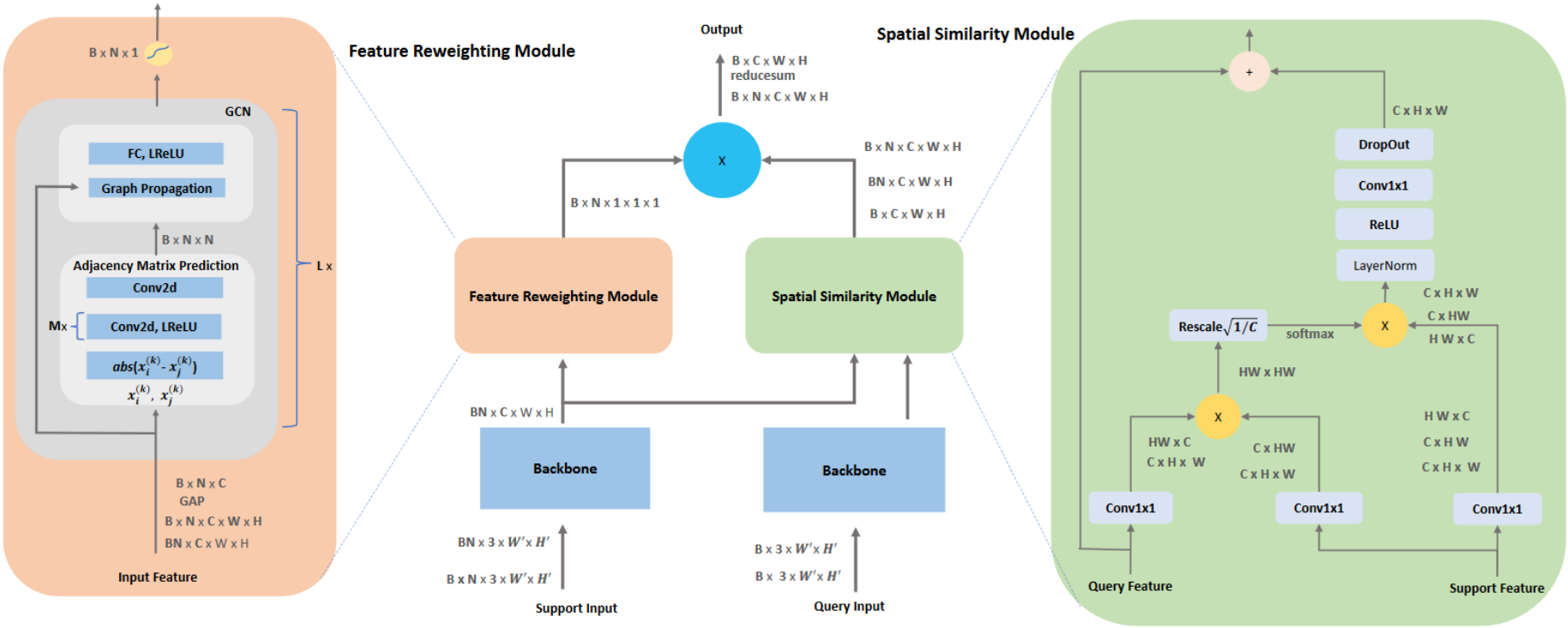
Method

Overview



Method

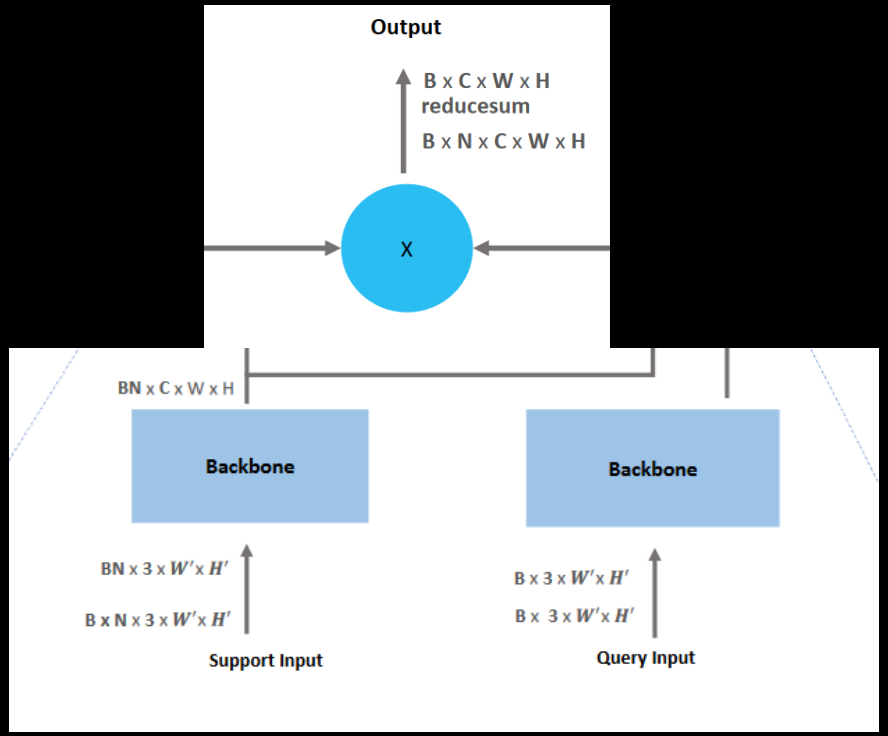
Overview



Backbone & Final Detection

- VGG feature extraction
- Get Query & Support features

- Multi scale box detection
  - Like SSD



Global Average Pooling: Baseline

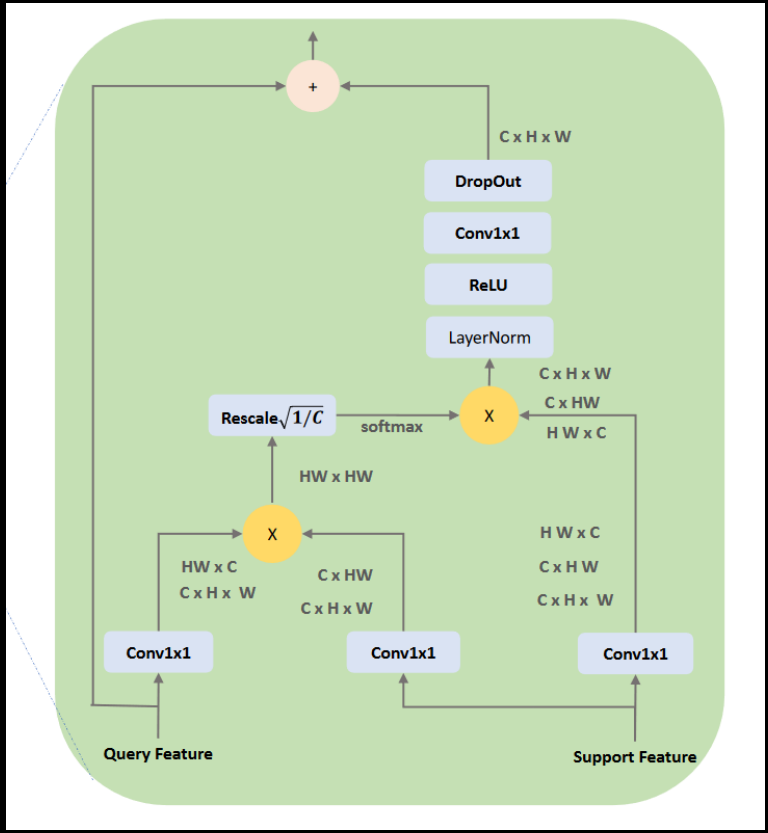
- Pool features from baseline
  - q_i - from query image
  - S_i - from support images
- Average features together

$$\phi(q_i, S_i) = q_i + \frac{1}{N}\sum_{j=1}^{N} \mathrm{GAP}(S_i^{(j)})$$

Method

Spatial Similarity Module

- Find spatial support between query and support images
- Image wise similarity (not global)
- Convolutions & Matrix mult
- Concat with q_i again at end



Method

Feature Reweighting Module

- Some support images may be more useful than others
- More salient features may not be as important
- Uses Graph Convolutional Network

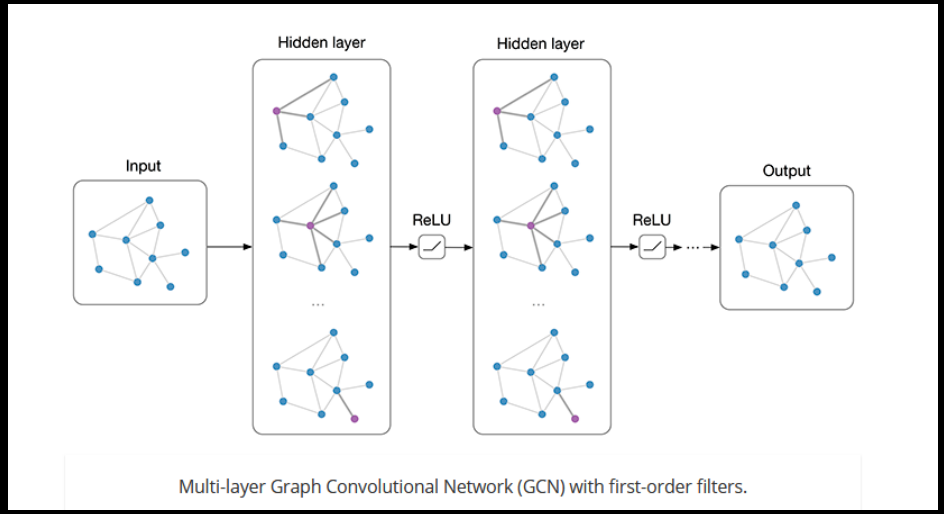




Method: Feature Reweighting Module

Graph Convolutional Networks

- Useful for graph datasets
  - Social networks, 3d skeletons, protein-interaction networks
- Can model other problems as graph problems
  - Eg. Image clustering

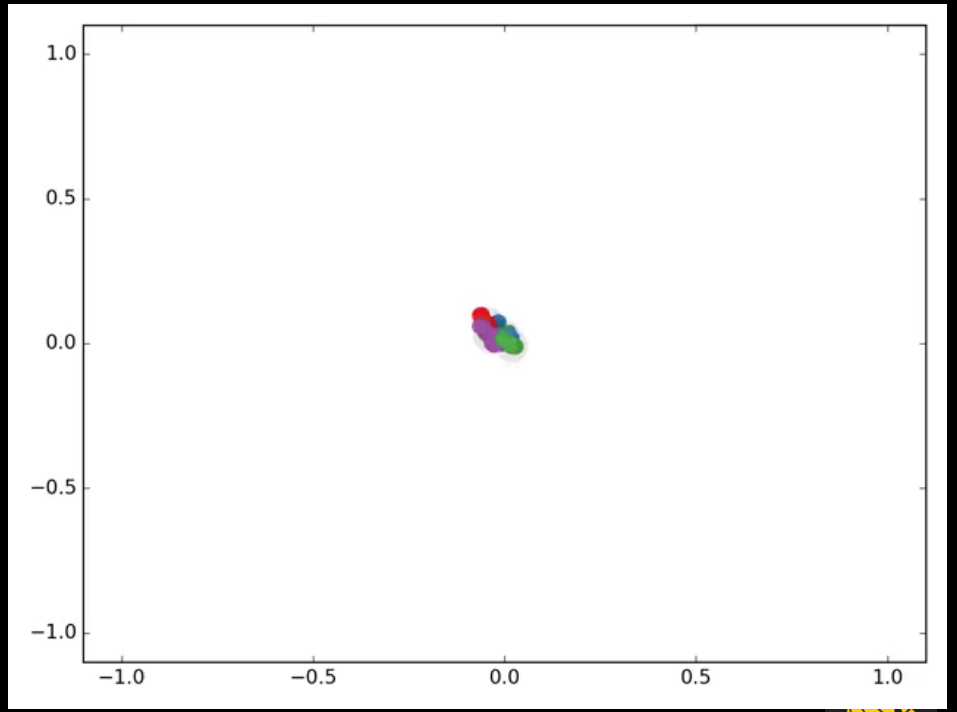


Multi-layer Graph Convolutional Network (GCN) with first-order filters.

Method: Feature Reweighting Module

Graph Convolutional Networks

- Nodes represent entities (images)
  - Each Node has feature vector
- Edges are relationships
  - Random or initialized
- Start with few target labels
- Subsequent layers relax adjacent edges



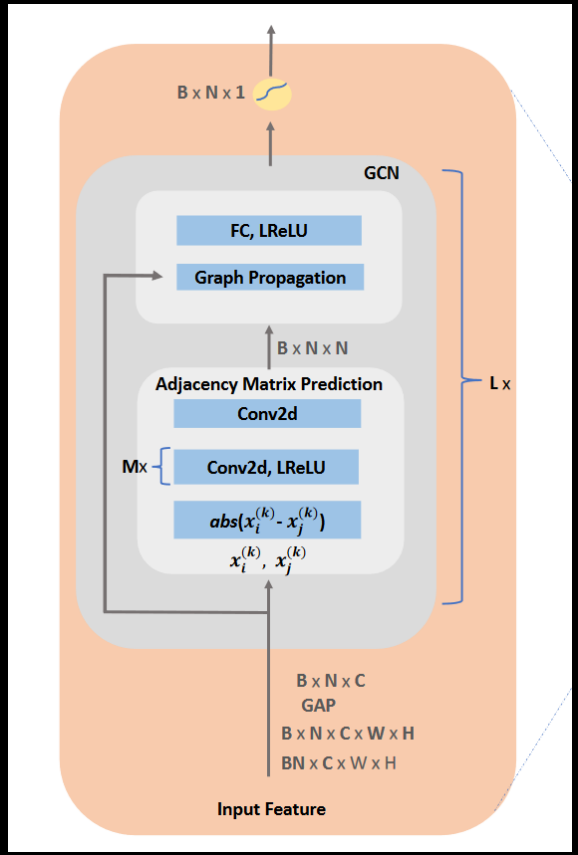
https://tkipf.github.io/graph-convolutional-networks/

UCF

Method

Feature Reweighting Module

- Uses GCN to reweight support images
- Predict adjacency matrix
- Graph convolutions to find important image clusters



Training

- Just one box per sample
  - Support images don’t require bounding boxes
- Losses:
  - Bounding box regression
  - Cross entropy classification

$$L(x, c, l, g) = \frac{1}{BD} \sum_{i,j=1}^{B,D} (\boldsymbol{bce}(c_{ij}, x_{ij}) + \ell_1^s(l_{ij}, g_{ij})), \quad (12)$$

UCF

Results



Experimental Setup

- Datasets modified for few-shot learning
- Image datasets are divided in half
  - One for training, other for testing
  - Then swap
- CL-VID used for testing only
  - Evaluate on each frame

UCF

Experimental Setup

- Datasets modified for few-shot learning
- Image datasets are divided in half
  - One for training, other for testing
  - Then swap
- CL-VID used for testing only
  - Evaluate on each frame

- Pascal VOC → CL-VOC
  - 20 class → 2x 10 class
- MS-COCO → CL-COCO
  - 80 → 2x 40
- ImageNet VID → CL-VID
  - Train on CL-VOC-12
  - 30 class, 3,862 clips

Ablation

- Spatial similarity
  - Local over global
- Feature reweighting
  - Catches and removes outliers

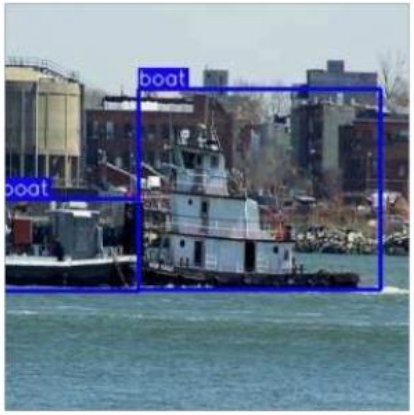
Table 2. **Ablation of spatial similarity and feature reweighting.** The metric is mean average precision (%). As we adopt spatial similarity (SSM) and feature reweighting (FRM) the accuracy gradually increases over a simple global average pooling (GAP), indicating the effectiveness of our proposed modules.

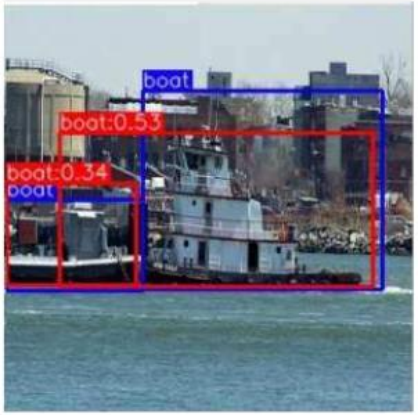
dataset	GAP	SSM	FRM	Group 1	Group 2	mean
CL-VOC-07	✓			55.17	52.18	53.67
		✓		56.12	55.52	55.82
		✓	✓	**57.17**	**56.45**	**56.82**
CL-VOC-12	✓			53.55	54.61	54.08
		✓		54.04	57.39	55.71
		✓	✓	**55.11**	**58.62**	**56.86**
CL-COCO	✓			13.37	6.50	9.94
		✓		18.50	7.70	13.10
		✓	✓	**18.62**	**8.20**	**13.40**

UCF





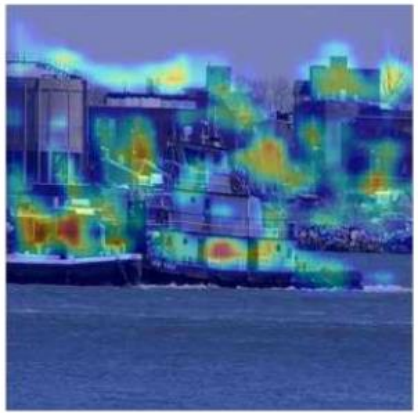

boat
boat


boat
boat:0.53
boat:0.34
boat

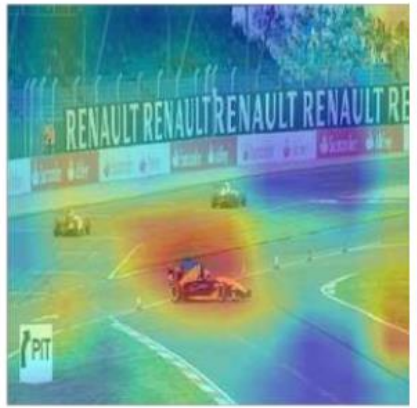

car:0.55
RENAULT RENAULT RENAULT RENAULT RE
car
PIT


RENAULT RENAULT RENAULT RENAULT RE
car:0.34
PIT



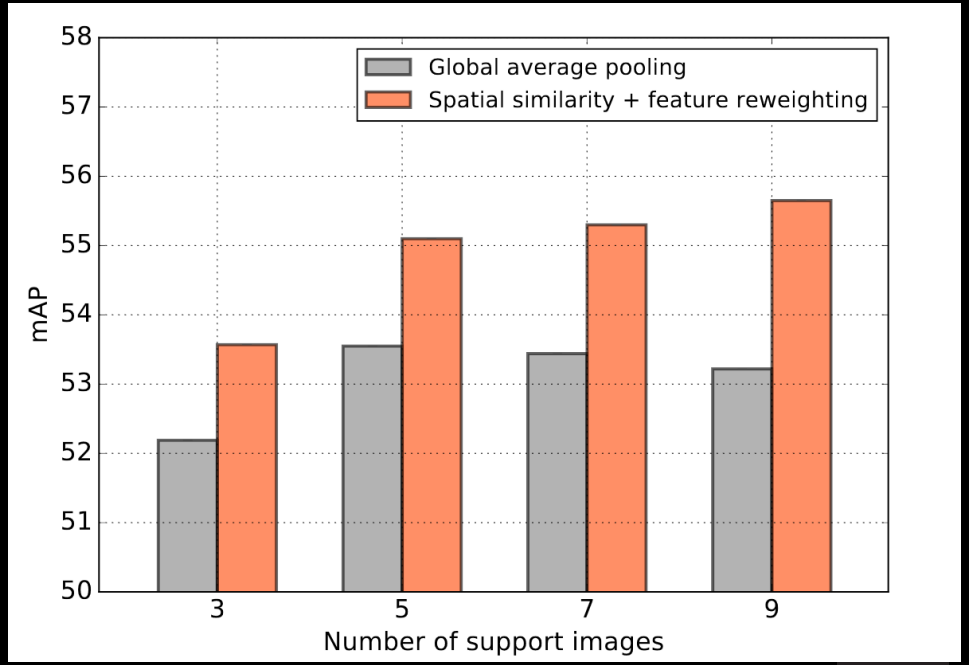



RENAULT RENAULT RENAULT RENAULT RE
PIT


RENAULT RENAULT RENAULT RENAULT RE
PIT

Effect of Support Images

- Peak precision reached early
- Gap w/ baseline widens



UCF

Results

Effect of Object Size

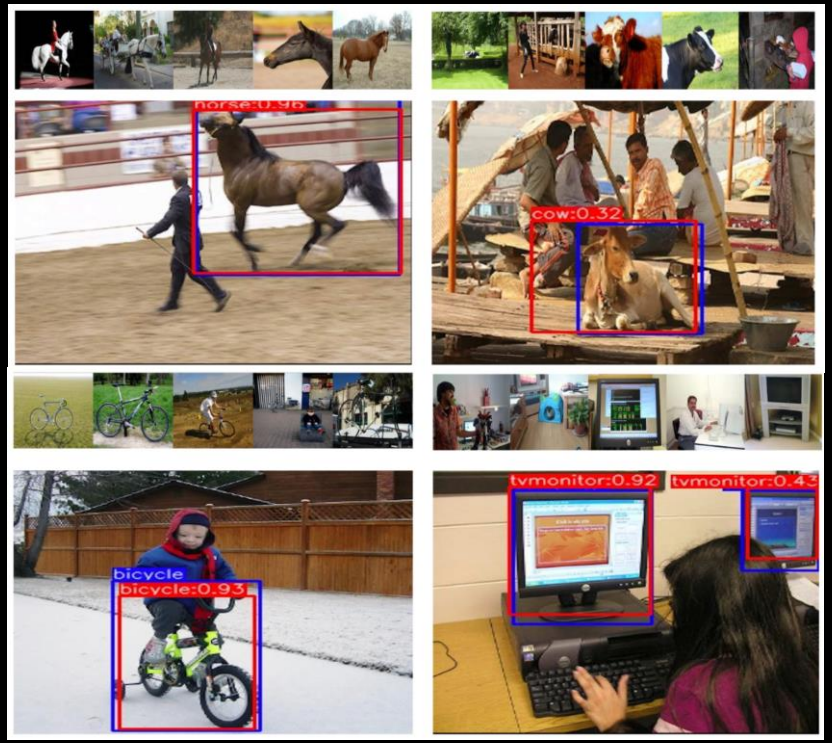
Table 3. **Effect of object size** on CL-VOC-12 in mAP(%). Compared to a global average pooling (GAP) baseline, our spatial similarity (SSM) and feature reweighting modules (FRM) especially improves for medium-sized objects.

method	small	medium	large
GAP	9.60	10.87	28.64
SSM, FRM	**10.99**	**13.85**	**29.24**



UCF

Success and Failure Cases

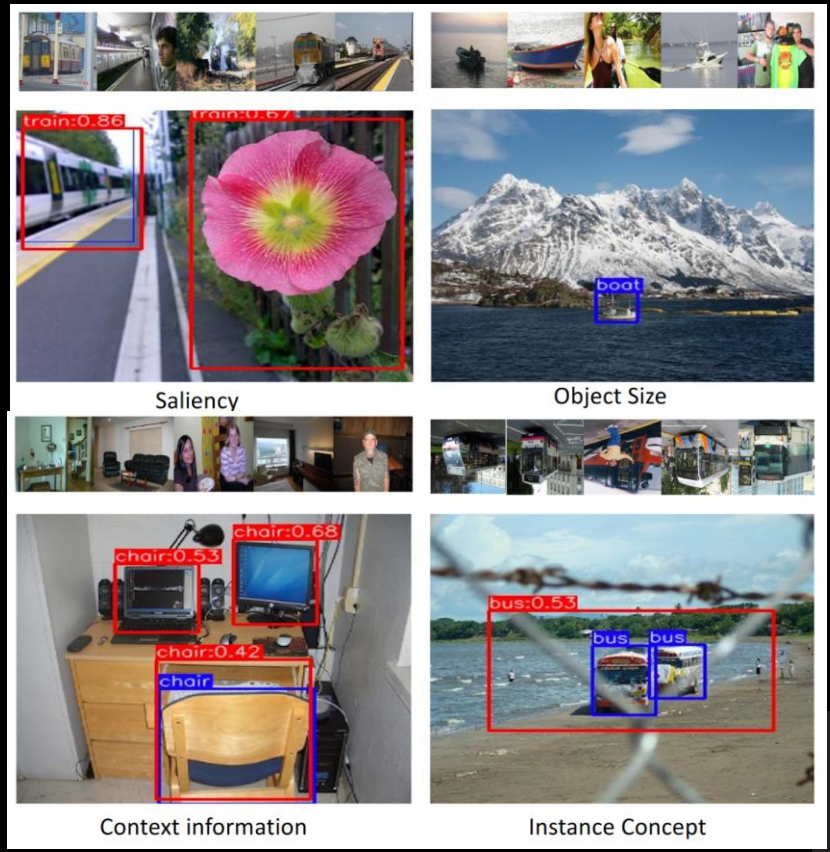


- Queries with multiple objects
- Multiple instances
- Corruption & bad examples

UCF

Success and Failure Cases

1. Saliency
2. Object Size
3. Context Information
4. Instance Boundaries



UCF

Comparative Evaluation

Table 4. **Comparative evaluation** on CL-VOC-07, CL-VOC-12, and CL-COCO. Across all datasets, our approach outperforms the center box and Region Proposal Network (RPN) [39] baselines by a large margin. Furthermore, our Spatial Similarity (SSM) and Feature Reweighting (FRM) modules are preferred over the ConvLSTM of HU *et al*. [24] for few-shot common-localization.

	CL-VOC-07			**CL-VOC-12**			**CL-COCO**			**CL-VID**
	group 1	group 2	mean	group 1	group 2	mean	group 1	group 2	mean	mean
Center box	5.28	4.05	4.67	6.56	6.90	6.73	0.54	0.51	0.53	-
RPN [39]	18.72	15.13	16.93	20.23	18.54	19.39	3.20	1.93	2.57	-
Contrastive RPN [19]	18.89	15.33	17.21	19.21	18.53	18.87	3.54	2.01	2.77	-
This paper										
w/ ConvLSTM [24]	53.46	54.90	54.18	51.34	57.42	54.38	16.37	6.72	11.54	53.02
w/ SSM and FRM	**57.17**	**56.10**	**56.64**	**55.12**	**58.62**	**56.87**	**18.62**	**7.47**	**13.04**	**54.40**

UCF

Summary and Conclusion

- Few-shot common-localization
- Cheap bounding boxes
- Spatial similarity module
- Feature reweighting module
- Beats baselines by large margin

UCF

Appendix: Useful Links

- Paper Website: http://taohu.me/SILCO/
- Supplemental: http://openaccess.thecvf.com/content_ICCV_2019/supplemental/Hu_SILCO_Show_a_ICCV_2019_supplemental.pdf
- Graph Conv Nets: https://tkipf.github.io/graph-convolutional-networks/
- Few Shot w/ Graph Nets: https://arxiv.org/pdf/1711.04043.pdf

UCF


UNIVERSITY OF CENTRAL FLORIDA
1963
UCF