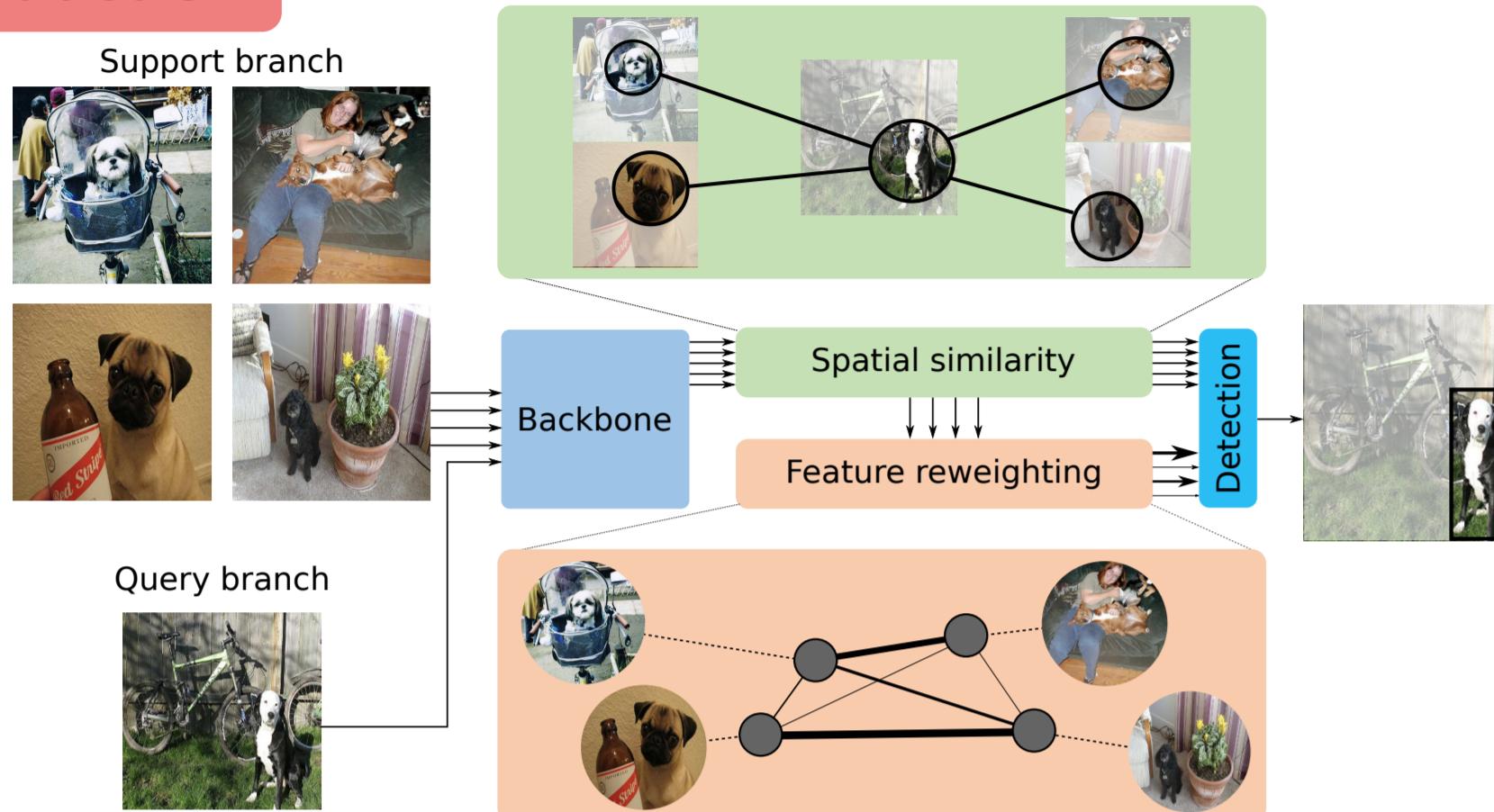


Introduction



We propose a new task named few-shot common-localization. Given a few weakly-supervised support images, we aim to localize the common object in the query image without any box annotation.

Approach

$$\tilde{q}_i = \phi(q_i, S_i), f(q, S) = \text{DET}(\text{CONCAT}_{i \in S}(\tilde{q}_i))$$

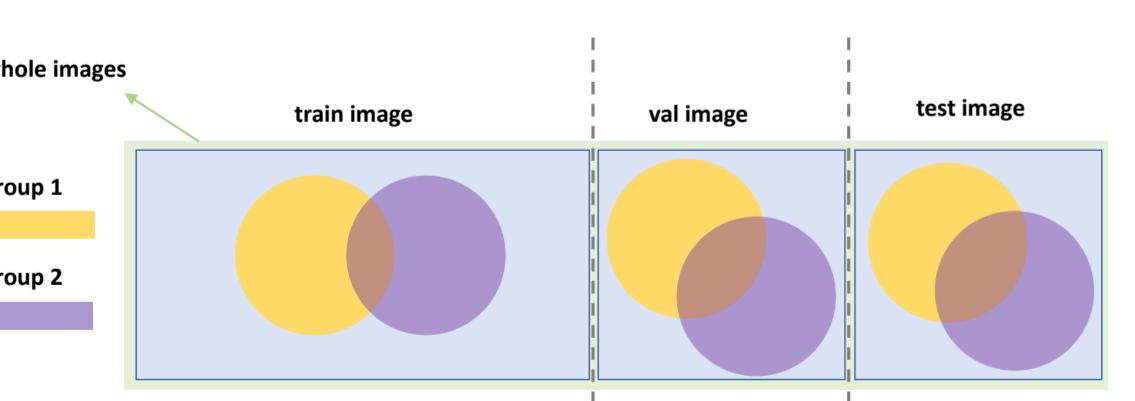
$$\phi(q_i, S_i) = \text{RS}(\text{CONCAT}_{j=1}^N (\text{SSM}_{im}^j(q_i, S_i)) \otimes \text{FRM}(S_i))$$

SSM: Spatial Similarity Module that searches the spatial commonality among the given images.

FRM: Feature Reweighting Module to balance the influence of different support images through graph convolutional networks.

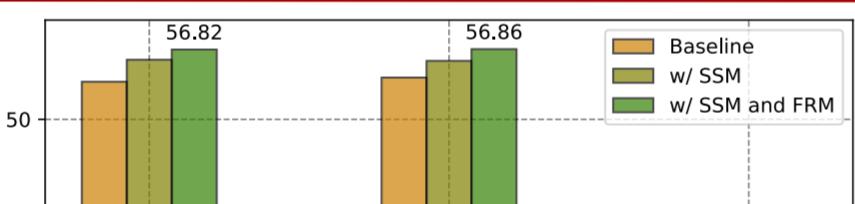
Result

CL dataset originates from Pascal VOC12 and MS COCO, the splitting principle is shown below. Our labels are separated between different images.

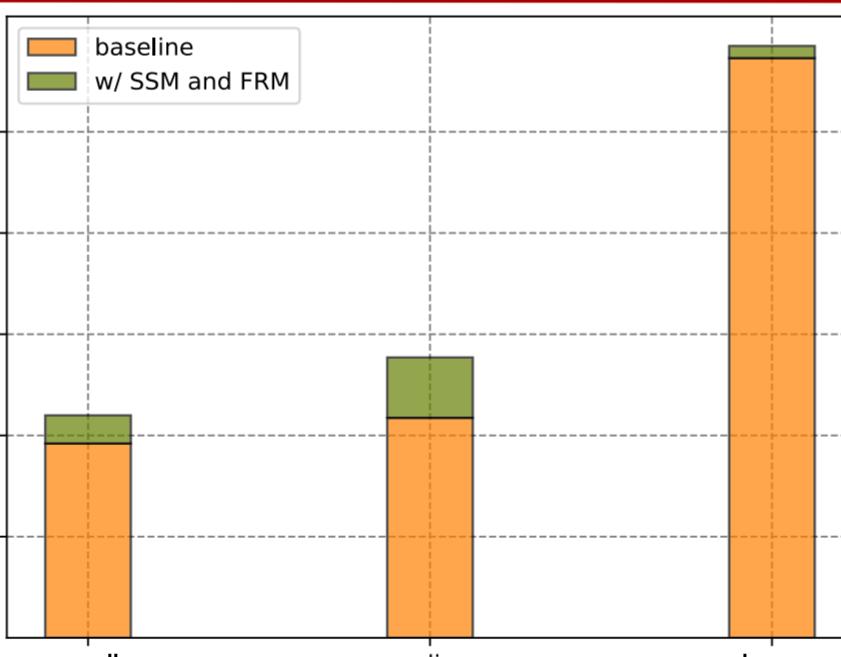


dataset	train	val	test
CL-VOC07	2501	1010	1500
CL-VOC12	5717	2623	3200
CL-COCO	62783	20000	40504

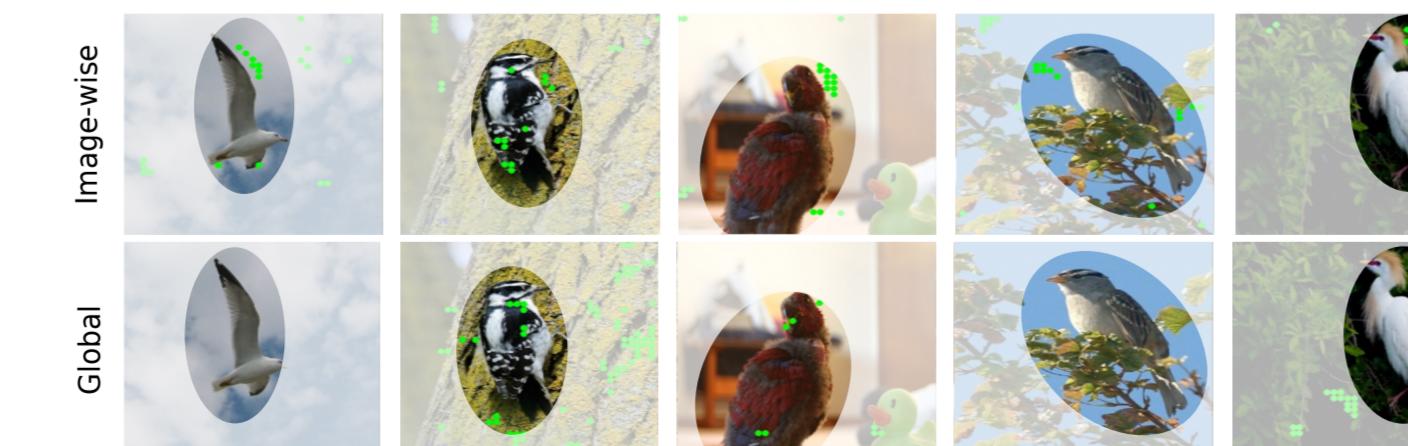
CL dataset statistics.



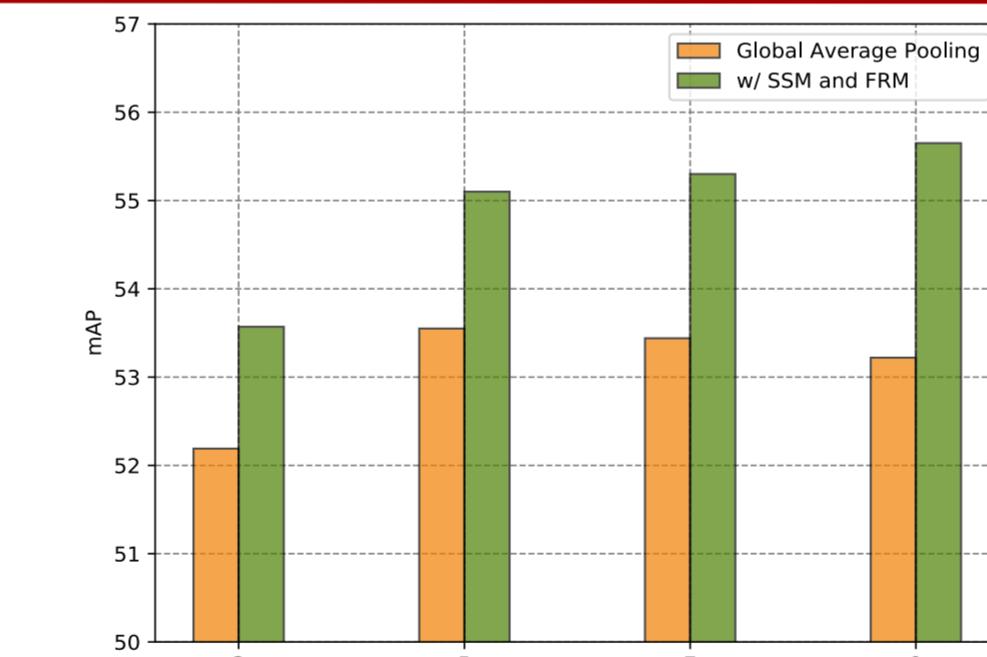
Module ablation: effectiveness of SSM and FRM.



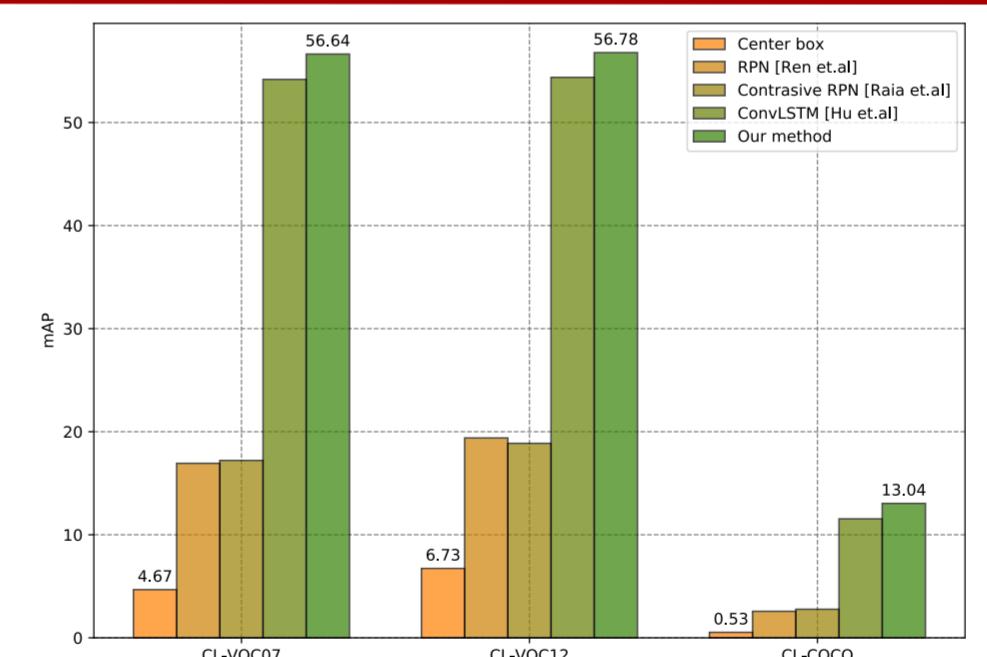
Object size ablation: most gain from the medium size objects.



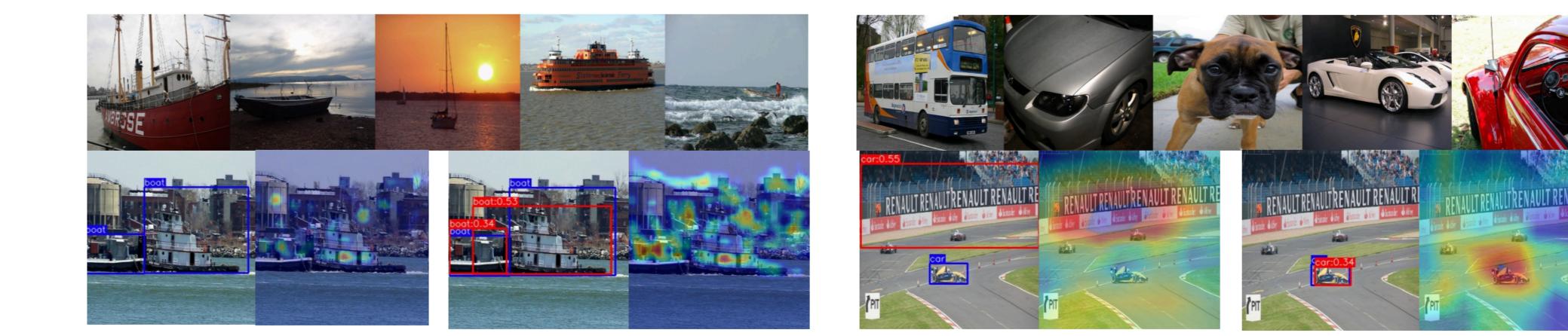
The left is the query image, the top, bottom images are image-wise similarity visualization and global similarity visualization respectively.



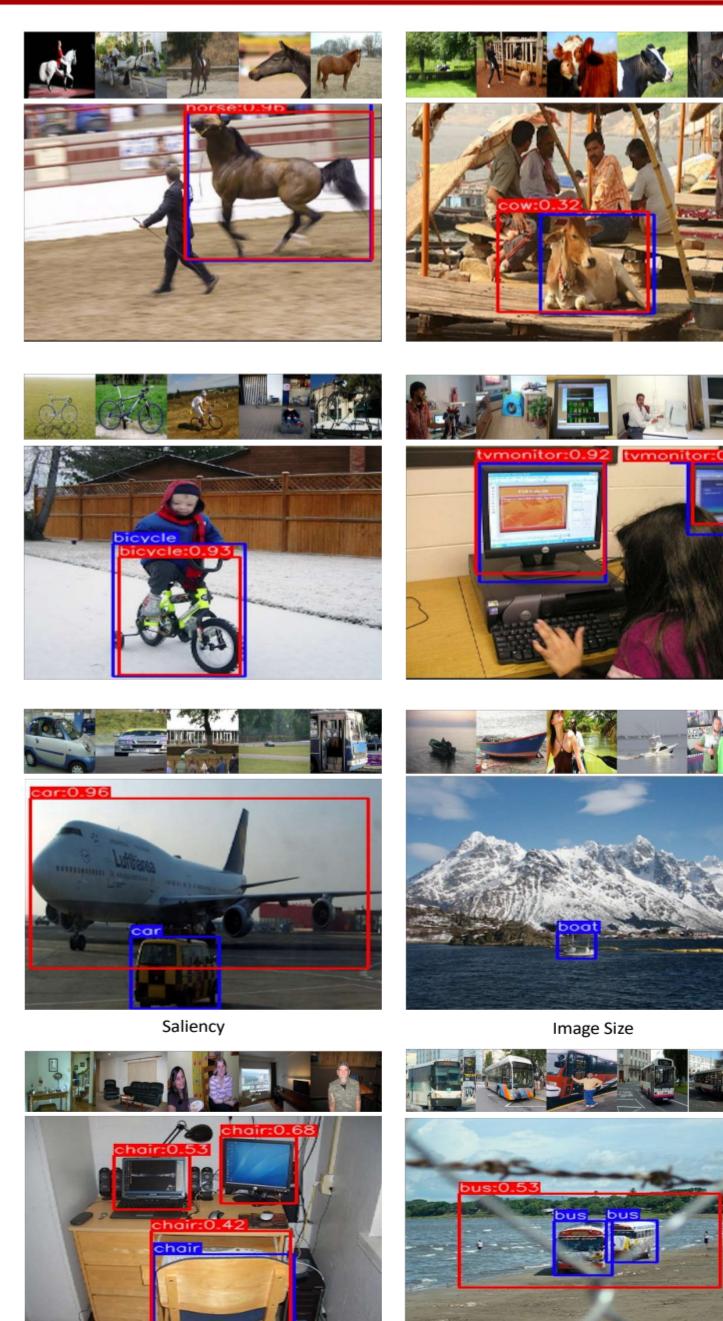
More gain as support set size increases.



Comparative evaluation on three datasets.

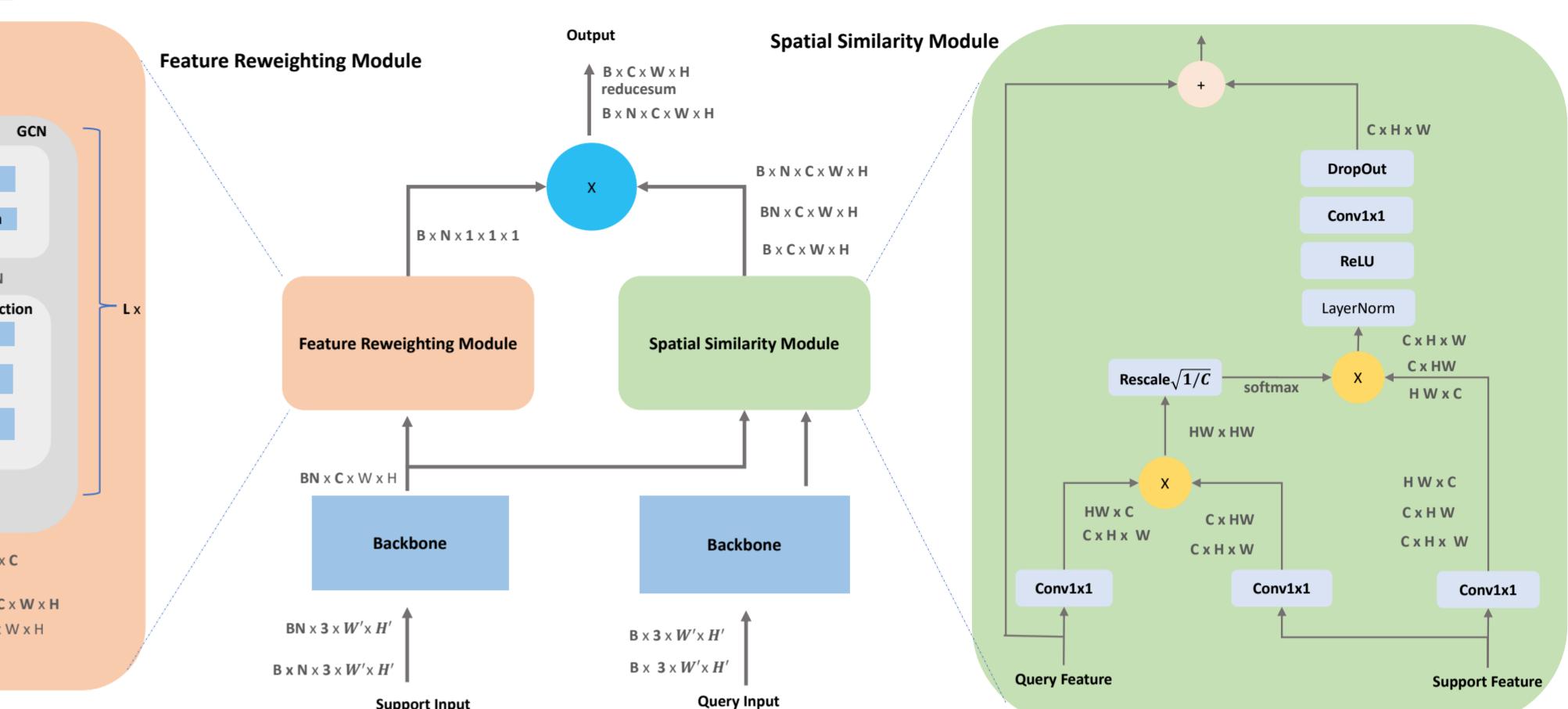


The heatmaps show feature reweighting better highlight common object.



Success and failure.

Architecture



Overview of the spatial similarity module, feature reweighting module, and the aggregation. L denotes the number of GCN blocks. M denotes the number of Conv2d-LReLU combinations. Graph Propagation means multiplication between vertex feature and graph adjacency matrix.

Code



taohu620@gmail.com

Looking for internship!