Restoring Negative Information in Few-Shot Object Detection

CODE: https://github.com/yang-yk/NP-RepMet We are **HIRING INTERN**! Please send your resume to fawe@microsoft.com

Yukuan Yang (Tsinghua university)
Fangyun Wei (Microsoft Research Asia)
Miaojing Shi (King's College London)
Guoqi Li(Tsinghua university)

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

Machine Learning VS Human Learning:

➢Machine learning: learning from *many samples*
➢Human learning: learning from *few samples*

Few-shot Learning:

Inspired by human learning, Few-shot Learning is defined as following. Using **prior knowledge**, Few-shot Learning can rapidly generalize to **new tasks** containing **only a few samples** with supervised information [1].

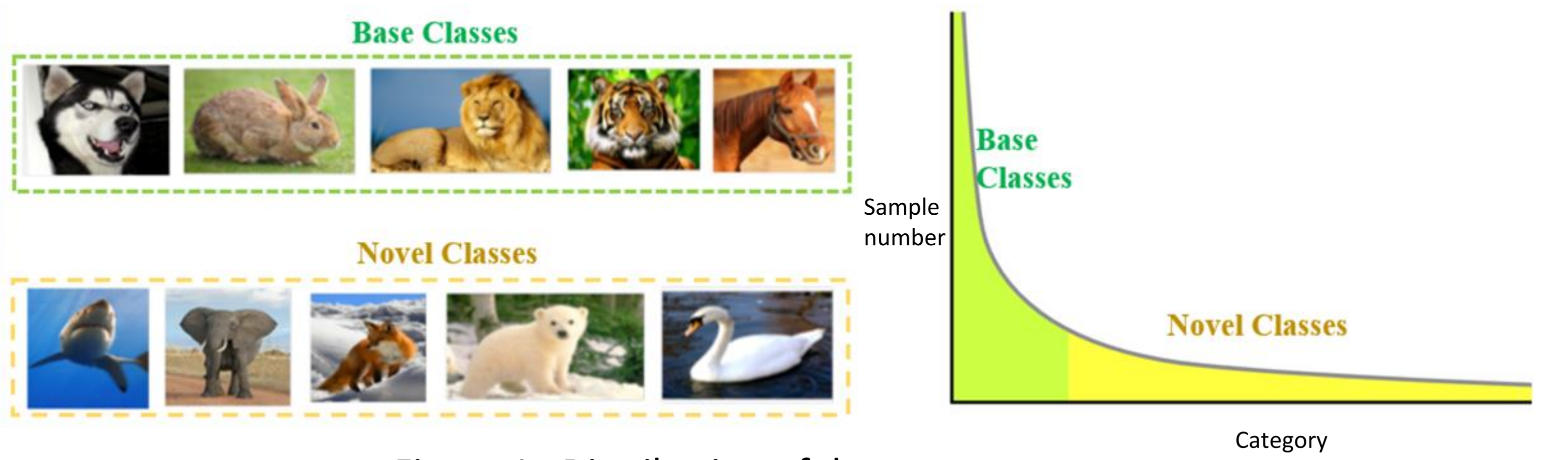


Figure 1. Distribution of dataset

- Base classes: **sufficient** samples and labels (*prior knowledge*)
- Novel classes: **limited** samples and labels (*weak supervised information*)
- New tasks: visual recognition (classification, detection, segmentation, etc.) on **novel classes** (*generalize to new tasks*)

Few-shot Detection:

Meta R-CNN [2]: *Extract positive information* of novel class through pre-trained backbone network.

RepMet [3]: *Learn positive information* of novel class through metric learning.

◆ Positive (foreground) information is extracted and utilized from support images.
◆ **Negative information** in support images is ignored in few-shot object detection.

NP-RepMet

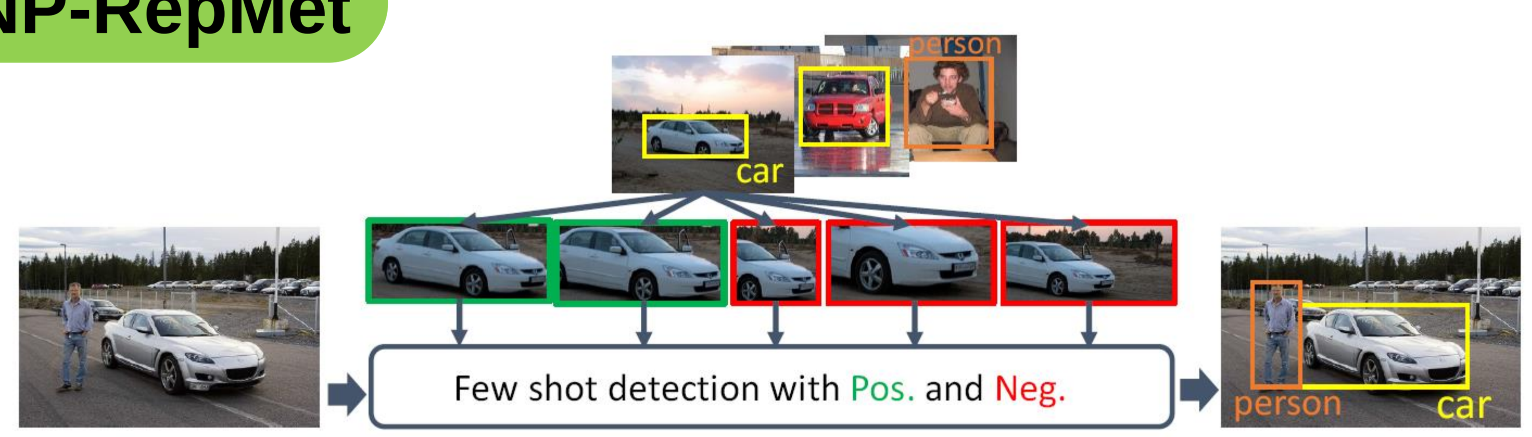


Figure 2. Restoring negative information in few-shot object detection.

How to explore the limited information obtained from novel class support set?
- Utilize **both positive and negative** information for better few-shot detection performance.

How to distinguish positive and negative information?
- Positive information: Proposals with **IoU > 0.7**
- Negative information: Proposals with **0.2< IoU <0.3**

How to combine positive and negative information in few-shot detection?
- Through **NP-RepMet** framework.

NP-RepMet Training:

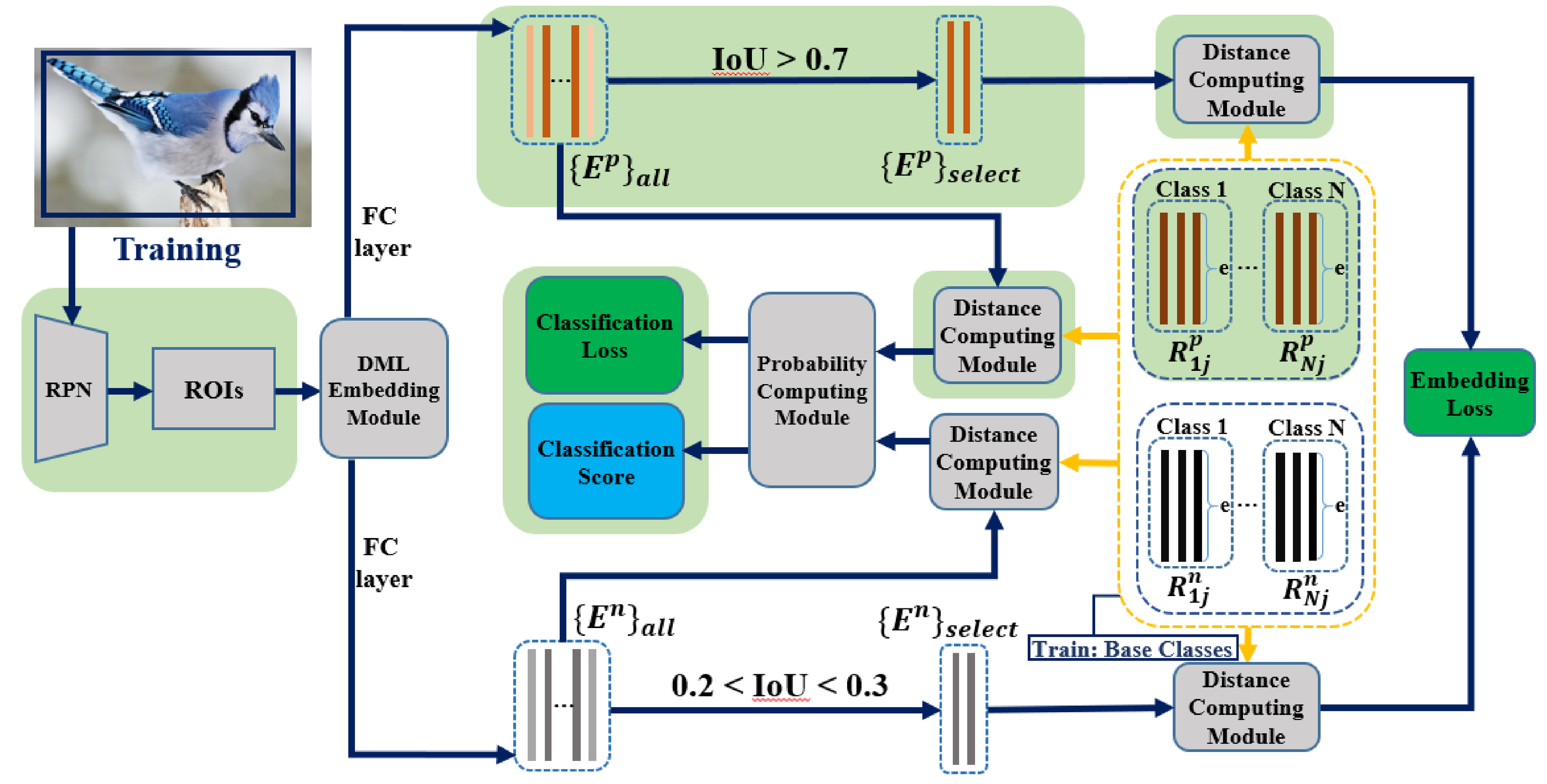


Figure 3. NP-RepMet training.

Triplet Embedding Loss:

If P is a positive proposal:

$$L(E^p,P)=|\min_j d(E^p,R^p_{i^*j})-\frac{1}{2}(\min_j d(E^p,R^n_{i^*j})+\min_{j,i\neq i^*} d(E^p,R^p_{ij}))+\alpha|_+$$

- **Minimize** the distance from positive embedding vector E^P to its closest **positive representative $R^p_{i^*j}$ of the same class;**
- **Maximize** the distance from positive embedding vector E^P to its closest **negative representative $R^n_{i^*j}$of the same class;**
- **Maximize** the distance from positive embedding vector E^P to its closest **positive representative $R^p_{i,j}(i\neq i^*)$of different classes**.

If P is a negative proposal:

$$L(E^n,P)=|\min_j d(E^n,R^n_{i^*j})-\frac{1}{2}(\min_j d(E^n,R^p_{i^*j})+\min_{j,i\neq i^*} d(E^n,R^n_{ij}))+\alpha|_+$$

- **Minimize** the distance from negative embedding vector E^n to its closest **negative representative $R^n_{i^*j}$ of the same class;**
- **Maximize** the distance from negative embedding vector E^n to its closest **positive representative $R^p_{i^*j}$of the same class;**
- **Maximize** the distance from negative embedding vector E^n to its closest **negative representative $R^n_{ij}(i\neq i^*)$of different classes.**

NP-RepMet Test:

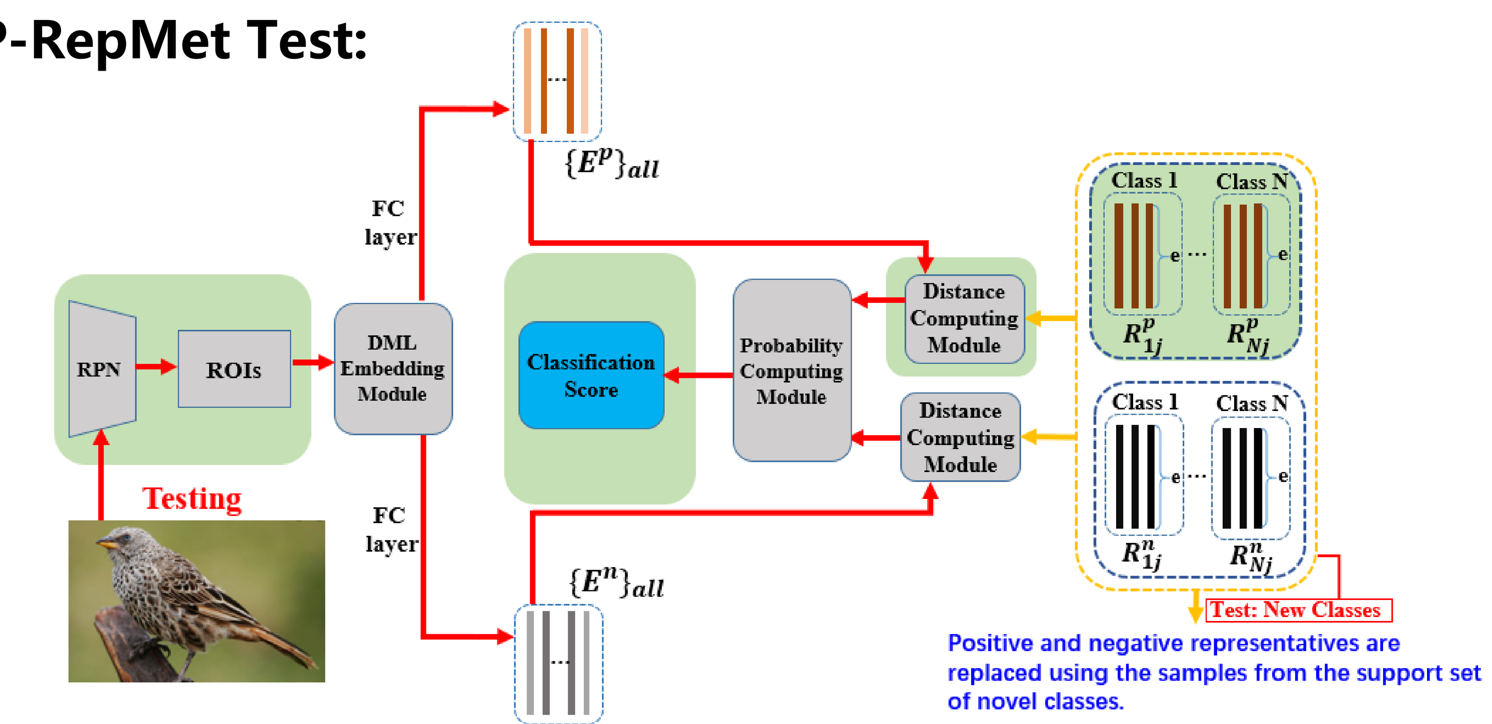


Figure 4. NP-RepMet test.

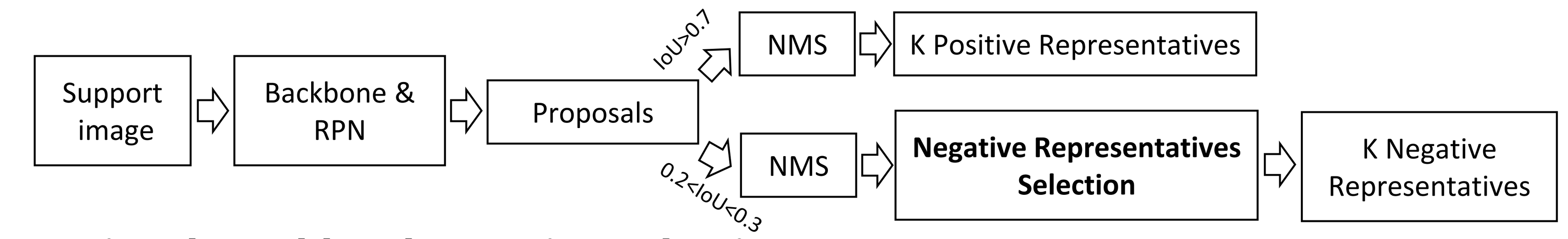
Combine Positive and Negative Information:

In the test process, a proposal generates two embedding vectors: **positive vector E^P and negative vector E^n**. For classification, we have

$$p_i(E^p,E^n)\propto\exp\left(-\frac{\min_j d(E^p,R^p_{ij})-\beta\min_j d(E^n,R^n_{ij})+2\beta}{2\sigma^2}\right)$$

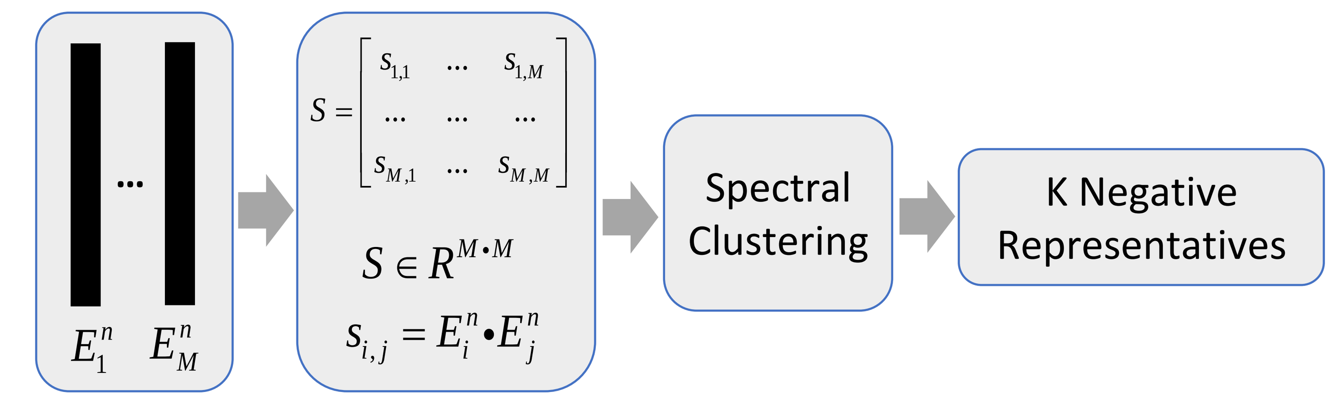
Negative Representatives Selection:

After NMS, the number of negative proposals is much bigger than that of positive proposals.



Clustering-based hard negative selection

Select the most **diverse** ones from negative proposal candidates.



Results

ImageNet-LOC:

NP-RepMet results compared with RepMet under the same setting.

11.6% mAP improvement in 5-way 1-shot detection

Dataset	Method	1-shot	5-shot	10-shot
ImageNet-LOC	RepMet	56.9	68.8	71.5
(214 unseen animal classes)	Ours	**68.5**	**75.0**	**76.3**
ImageNet-LOC	RepMet	86.0	90.2	90.5
(100 seen animal classes)	Ours	**93.7**	**94.0**	**95.3**

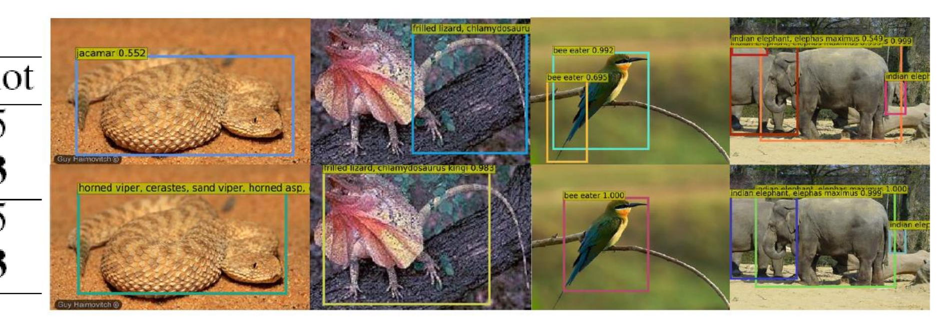


Figure 5 .Detection results of RepMet (Top) and NP-RepMet (Bottom).

Pascal VOC:

NP-RepMet results compared with Meta R-CNN under the same setting.

17.9% mAP improvement in 5-way 1-shot detection

Method/Shot	novel set 1					novel set 2					novel set 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
YOLO-FR [10]	14.8	15.5	26.7	33.9	47.2	15.7	15.3	22.7	30.1	39.2	19.2	21.7	25.7	40.6	41.3
Meta-Det [12]	18.9	20.6	30.2	36.8	49.6	21.8	23.1	27.8	31.7	43.0	20.6	23.9	29.4	43.9	44.1
Meta R-CNN [11]	19.9	25.5	35.0	45.7	**51.5**	10.4	19.4	29.6	34.8	45.4	14.3	18.2	27.5	41.2	**48.1**
RepMet [1]	26.1	32.9	34.4	38.6	41.3	17.2	22.1	23.4	28.3	35.8	27.5	31.1	31.5	34.4	37.2
Ours	**37.8**	**40.3**	**41.7**	**47.3**	49.4	**41.6**	**43.0**	**43.4**	**47.4**	**49.1**	**33.3**	**38.0**	**39.8**	**41.5**	44.8

T-SNE visualization:

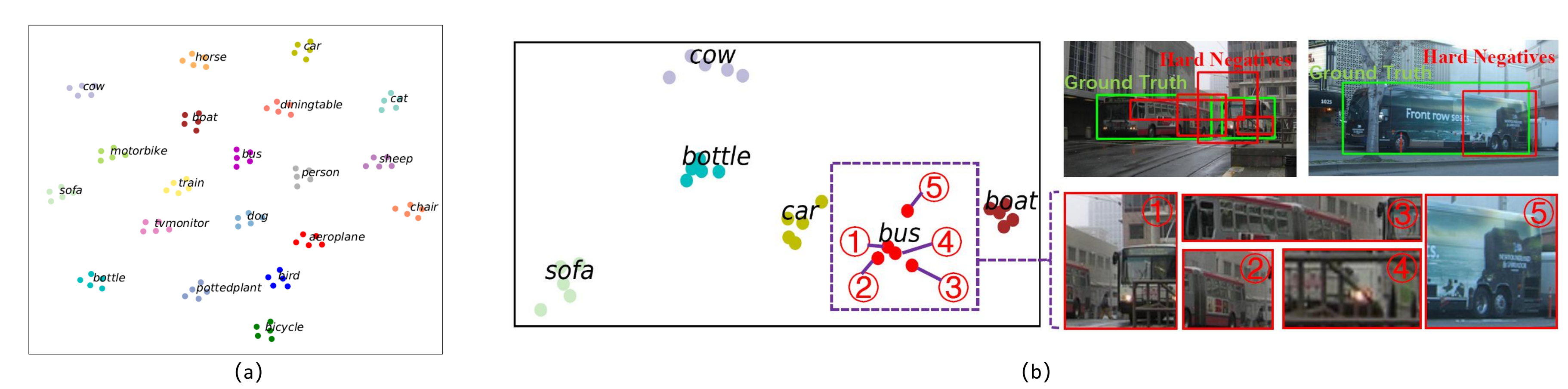


Figure 6. T-SNE visualization (a) T-SNE visualization of positive representatives (b) T-SNE visualization of negative representatives and corresponding proposals.

References

[1] Wang Y, Yao Q, Kwok J T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. ACM Computing Surveys (CSUR), 2020, 53(3): 1-34.

[2] Yan X, Chen Z, Xu A, et al. Meta r-cnn: Towards general solver for instance-level low-shot learning, CVPR 2019.

[3] Karlinsky L, Shtok J, et al. Repmet: Representative-based metric learning for classification and few-shot object detection, CVPR 2019.