Few-shot Learning: Towards localization and classification of objects

Jaswinder singh[1], B.K.Sharma[2]

Abstract
When we have dataset with large number of labelled examples it is easy to perform object detection task but, rare object detection from a few examples is a new problem. Meta-learning has been shown to be a promising strategy in the past. However, fine-tuning strategies have received little attention. We discovered that fine-tuning the last layer of detector is a critical task in few-shot object detection. On current benchmarks, such a basic strategy outperforms meta-learning approaches by about 4 to 16 points and sometimes the accuracy is doubled when compared to existing methodologies. However, current benchmarks are frequently unreliable because of the significant variance in the few samples. To generate consistent comparisons, we change the evaluation processes by choosing various sets of training examples. The model has been evaluated on three datasets: COCO, LVIS, and PASCAL VOC. Our fine-tuning approach amalgamated with the Ranking based loss function which can be used for both classification and localization is state-of-the-art.

Keywords: *Computer Vision, Few-shot Object detection, Meta-learning,*

Introduction

Recently, machine perception software has made great improvements. However, when in comparison to human visible systems, our capability to train versions that generalise to novel concepts with less amount labelled data is still lacking. With very little guidance, even a toddler may instantly recognise a new notion(Samuelson & Smith, 2005; Smith et al., 2002). Few-shot learning, or the ability to generalise from a small number of examples, has become a hot topic in the machine learning community.

Many(Finn et al., 2017; Gidaris & Komodakis, 2018; Vinyals et al., 2016) researchers have looked into meta-learning as a way to transfer knowledge from data-rich base classes to data-poor innovative classes. To run few-shot simulations on novel classes the researchers do sampling and form a base class during training(X. Wang et al., 2020). However, the majority of this research has been on fundamental picture categorization problems. Few-shot object detection, on the other hand, has gotten significantly less attention. Object detection, unlike picture classification, necessitates the model to not only distinguish object kinds but also to locate the targets among millions of possible regions. The overall complexity of the task is significantly increased by the addition of this new subtask(Kang et al., 2019)(Y. X. Wang et al., 2019). Several(Kang et al., 2019; Ren et al., 2017; Y. X. Wang et al., 2019; Yan et al., 2019) researchers have attempted to solve the object detection problem with few labelled

[1] PhD Scholar., Dr. A.P.J. Abdul kalam University, Department of Computer Science,w.s.jaswinder@gmail.com

[2] Dr., NITRA Technical campus, Ghaziabad, India, drbksharma@nitratextile.org

data, the few-shot learning problem generally is solved by attaching a meta learner which uses Meta-learning methods for classification and then a regressor to detect the underlying object networks.

However, present evaluation techniques are statistically unreliable, and the accuracy of baseline approaches on few-object identification, particularly easy fine-tuning, is inconsistent throughout the literature(A. Li et al., 2020).

We propose a novel approach to detect object using few labelled data i.e. our model is highly adaptive to limited number of new samples. We circumspectly analyze calibrating based strategies that are considered to fail to meet expectations in the prior works(Kang et al., 2019; Y. X. Wang et al., 2019). We are attempting to bridge the divide between picture classification and object detection. Detection, unlike image classification, necessitates the identification of (potentially many) objects inside an image. We focus on the preparation standard just as the example level capacity standardization of the specific article identifiers inside model plan in addition to preparing contingent upon calibrating. We receive the two-stage preparing plan for calibrating. We all first train the total article indicator, this sort of as Faster R-CNN(Ren et al., 2017), around the data abundant base courses, after which just adjust the last layers related with the specific identifier upon little adjusted instructing masterminded involving every establishment and book courses while freezing the specific different rules related with the plan(T. Wang et al., 2019). All through the adjusting stage, we all present case level capacity standardization to the specific bundle classifier(Gidaris & Komodakis, 2018).

On the current PASCAL VOC [13] and COCO [14] standards, we have achieved state-of-the-art performance when compared to all previous available methods by 4 to 16 points. Our method can perform twice as well as previous superior state-of-the-art systems when training on a single (one-shot learning) shot example. Existing analytic methods have a number of flaws that prevent consistent model comparisons. Precision measurements have a lot of fluctuation, thus released comparisons aren't very accurate. Furthermore, past evaluations only reported detection accuracy on fresh courses and did not examine knowledge retention in foundation classes.

We're all working on new standards based on three datasets: COCO, LVIS [15], and PASCAL VOC. To get a steady accuracy evaluation and quantitatively examine the variations associated with distinct examination measures, we all sample distinct groupings of few-shot instruction instances for several runs from the tests. The new evaluation reports the mean accuracy on all lessons, referred to the generic few-shot learning setting within the few-shot category [8], as well as the typical accuracy on both the bottom and novel lessons. Our fine-tuning method creates entirely new states based on the artwork on the standards. On the difficult LVIS dataset, the two-stage training structure increases the typical detection precision associated with rare classes (less than 10 images) by approximately six points and typical classes (between two to 100 images) by approximately four points with all the minimal precision loss regarding the frequent lessons (on more compared to 100 images).

Meta-learning

Humans can recognise different objects with different dimensions even if they are provided very few examples, but it is a very difficult task in computer vision when the labelled data is limited. Meta-learning tries to achieve this goal when we have limited labelled data.

Researchers in(Berkeley et al., 2017; Ren et al., 2017) learn to fine-tune and purpose a great parameter setup that can adapt to the new tasks with just a few scholastic gradient upgrades. The use of parameter development during adaptation to novel tasks is another major line of meta-learning study. To produce final classifier values for the novelty classes, several researchers have proposed an attention-based weight generator. Some(Gidaris & Komodakis, 2018),(Kolmogorov & Rol, n.d.) researchers create task-aware pattern embeddings by producing feature layer settings. These methods have only been applied to single-shot picture classification and not to more difficult tasks such as object detection. Moreover, given the lack of a regular comparison of alternative methodologies, some researchers [11],[17],(Liu et al., 2020) are concerned about the results' dependability. Many prior efforts that apply meta-learning on few-shot image categorization turn out to be more beneficial than some basic fine-tuning-based algorithms, which have received little attention in the field. Due to rising network complexity and unclear implementation details, there are no common evaluation parameters for few-shot object detection task.

Metric-learning

Another area of study is(Vinyals et al., 2016) learning to compare, also known as metric learning. Even if we teach a child how a car looks, it is possible for them to recognise another car though it can have complete different feature sets like size, colour, and brand. This is how metric-learning works. So if a model is able to estimate the similarity between two different objects using distance metric, it should be able to generalise it into a novel category even when few labelled examples are provided. Recently, multiple researchers(Jiang et al., 2021; Karlinsky et al., 2019; X. Li et al., 2020) have used a cosine similarity-based classifier which reduces intraclass variance on the few-shot classification problem, resulting in better performance than several meta-learning-based techniques. To identify the categories of the region proposals, our method uses a cosine similarity classifier. However, rather than measuring distance at the image level, we focus on instance-level distance measurement.

Ranking based Object detection

Optimizing a ranking-based aim is an inspiring way for balancing classes(Kolmogorov & Rol, n.d.)(Chabot et al., 2019). However, because such objectives are discrete in relation to the scores, direct inclusion is difficult(Tan et al., 2019). The use of black-box solvers for an interpolated Average precision (AP) loss surface is one approach, however it only produced a minor speed boost. AP Loss provides an alternative technique, calculating gradients using an error-driven update method. Distributional ranking (DR) Loss, an alternative, uses Hinge Loss to establish a buffer between the positive and negative scores. Regardless of promising results, these kinds of methods are restricted to classification and depart localisation(Lv et al., 2021). As opposed, we all propose an individual, well-balanced, ranking-based loss to be able to train both divisions.

Object detection using fewer examples

Several initial meta-learning methods at the few object identification have been made. Many use feature re-weighting approaches with a meta learner that accepts the support pictures (i.e., a small number of labelled photos of the novel/base classes) and the boundary box annotations as inputs to a single-stage object detector (You only look once version2)(Oksuz et al., 2018) and a two-stage object detector (Faster R-CNN)(Ren et al., 2017). Another option is to utilise a weight prediction meta-model to estimate category-specific component attributes from a small number of instances while studying category-agnostic components from user defined class instances using a weight prediction meta-model(A. Li et al., 2019). Fine-tuning-based procedures are deemed baselines in all of these studies, with lower performance as compared to approaches using meta- learning. The model are trained using two different approaches (I) fine-tuning of model is done, when detector is trained is trained on both novel and base class(Girshick et al., 2014). (II) When the detector is trained only on base class and fine-tuned on novel and base class with balanced partition set(Xu et al., 2020). In our novel approach we have estimated that it is only sufficient to fine-tune the last layer of detector and rest layers can be ignored, the performance increase to almost two folds when trained using our approach and successfully outperformed all previous approaches of meta-learning.

Method

In this section we will be presenting our strategy based on two-stage fine-tuning approach for object detection. The faster R-CNN (FCRN), which itself is a two-stage detector is used as our basic detector. Figure 1 presents abstract of the approach which we are using, ResNet and VGG16 are used as the foundation for the region proposed network, and for feature extractor we are using a two-layer fully-connected sub-network. The feature acquisition component is denoted by F. The box predictor consists of a box classifier(C) which is used for categorizing types of objects and a regressor(R) which predicts the boundary coordinates. The R-CNN and region proposed network seem to be class-agnostic, which leads to the conclusion that the characteristics acquired from base class is likely to be carried over to a new class without any further parameter tuning. The division of feature representation and box prediction learning into two phases is a key feature of our approach.

Phase 1: Training the base model:

Step 1: Train box predictor and feature extractor on base class denoted by B_c.

Step 2: Use the loss function of [9].

Step 3: calculate joint loss

$$L = Lrpn + Lcls + Lloc \tag{1}$$

Where $Lcls$ is the cross entropy loss of classifier

$Lloc$Is smoothed for regressor

Step 4: Apply $Lrpn$ to output of RPN, which will differentiate foreground and background.

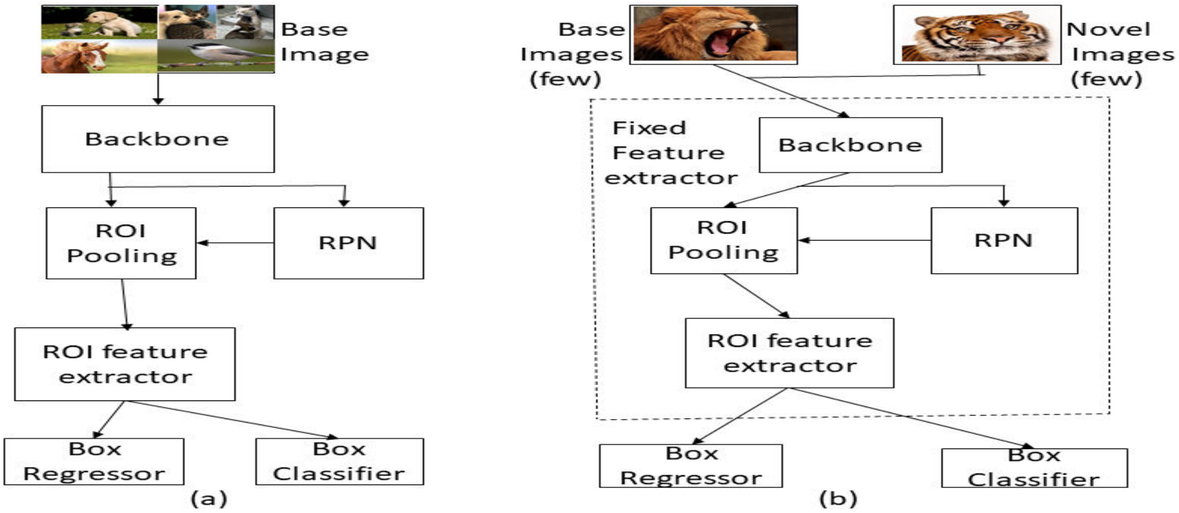

Figure 1: two-stage fine-tuning methodology is depicted in this diagram (TFA**).**

Phase 2: Fine-tuning

K (few) shots per class is extracted from the available samples and training class is designed. This includes both base and novel data. As demonstrated in the figure 2, the box prediction network is initialized with random weights and fine-tuning of the box classifier (C) and regressor (R) with only the last layer of detector is done. The equation 1 is used to apply the loss function but with lower learning rate. From the beginning we have observed 20 percent deduction in training rate of our model.

A meta-learner is used to obtain task-level Meta information which assist in the generalization of model on novel classes, through feature re-weighting. A two stage training strategy combined of meta-training and Meta fine-tuned approach is typically used with episodic learner.

Implementation details

We employ Resnet-101 with a Feature Pyramid Network as our backbone and FCRN as our base detector. The model is trained using batch approach with stochastic gradient descent with a batch size of 32 and weight decay factor is 0.0001. The momentum is fixed at 0.8. The learning rate was 0.01 and for few-shot fine tuning a rate of 0.0001 was used while training the base model.

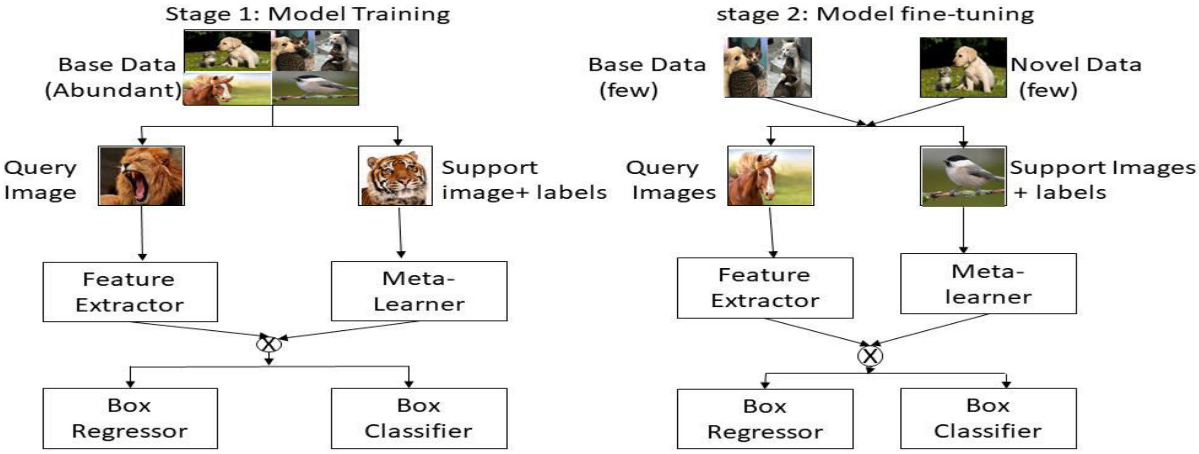

Figure 2: Fine-tuning model based on meta-based approach.

Baseline

We have compared our approach with previous meta-learning based approaches such as Meta-RCNN, FSRW and MetaDet, as well as the approach have been compared by other fine-tuning methods like FRCN when combined with YOLO, in which the base and novel approaches are trained in a single phase and the entire model (FCRN or YOLO with ft-full) is fine-tuned. In this the feature extractor and package predictor (C and R) both are fine-tuned. The FCRN or YOLO with ft- full is fine tuning the model with fewer labelled data is reported in (Kang et al., 2019)(Yan et al., 2019).

Existing Benchmark:

Model	**novel AP**		**novel AP^{75}**	
	10	30	10	30
FSRW	5.6	9.1	4.6	7.6
MetaDet	7.1	11.3	6.1	8.1
FRCN+ft+full	6.5	11.1	5.9	10.3
Meta R-CNN	8.7	12.4	6.6	10.8
FRCN+ft-full	8.9	12.6	9.3	14
TFA w/ fc	10	13.3	9.1	13.1
TFA w/ cos	10	13.6	9.2	13.3

Table 1: Comparison of results with existing benchmark on COCO dataset.

In the table 1 we have presented the benchmark in which we have compared FSRM(Kang et al., 2019), MetaDet(X. Wang et al., 2020), FRCN+ft+full(Yan et al., 2019) and Meta-R-CNN(Yan et al., 2019) with our reimplemented model of FRCN+ft+full, FC-based classifiers and cosine similarity based classifiers. Our model has consistently outperformed the existing benchmark.

Findings

In this section we have presented our research finding by implementing our novel approach on COCO and PASCAL VOC.

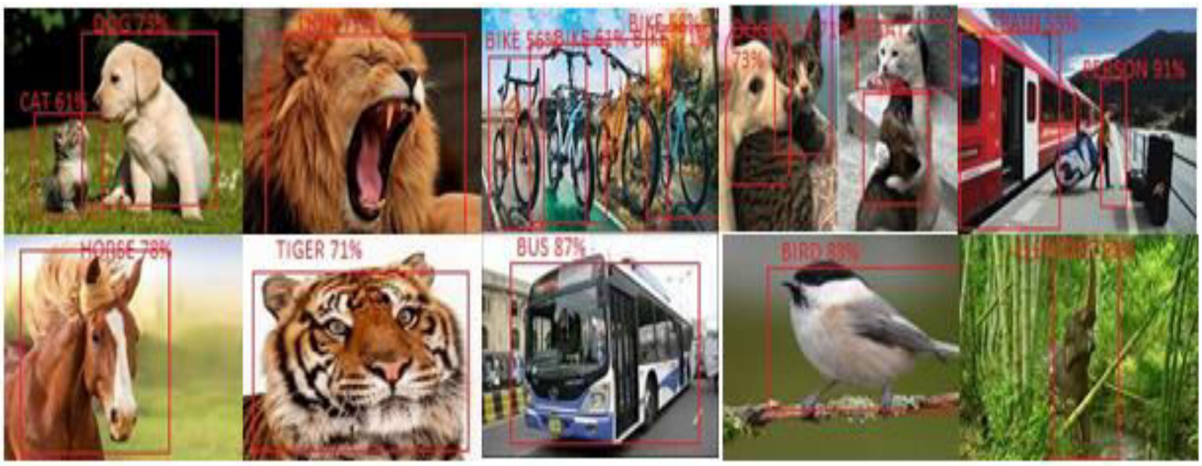

Figure 2: Object detection using our algorithm on PASCAL VOC and COCO dataset.

The figure 2 shows the result of our approach on PASCAL VOC (bus, train, bike) and COCO (cat, dog, horse, lion, tiger, elephant) dataset. The red boxes indicate the objects and their corresponding percentage. We are using a threshold value of 50%. In the cosine similarity we have initialized α to 25 and it has outperformed the previous existing methods on both COCO and PASCAL VOC in all our experiments.

Split	**# shots**	**Method**	**Base class**			**Novel class**			**Overall**		
			bAP	**bAP^{50}**	**bAP^{75}**	**nAP**	**nAP^{50}**	**nAP^{75}**	**AP**	**AP^{50}**	**AP^{75}**
Split 1	1	FSRW	34.1	62.9	32.6	8	14.2	7.9	26.6	50.8	26.5
			38.2	62.6	40.8	6	9.9	6.3	29.5	49.4	32.2
			48.7	77.1	53	12.2	22.9	11.6	39.6	63.5	43.2
			49.4	77.6	54.8	14.2	25.3	14.2	40.6	64.5	44.7
	3	FSRW FRCN+ft-full	33.9	61.8	32.7	13.2	24.2	12.6	28.7	52.2	27.7
			37.3	60.7	40.1	9.9	14.2	10.2	30.5	49.4	32.6
			47.8	75.8	52.2	18.9	33.24	18.2	40.5	65.5	43.8
			49.6	77.3	55	21.7	35.2	21.6	42.6	67.1	47
	5	FSRW FRCN+ft-full	32.4	60.5	93.35	74.68	29.4	15.4	29.5	52.4	27.5
			36.7	60.3	34.94	38.40	20.4	13.8	31.8	50.2	33.2
			47.6	77.4	30.92	50.74	40.2	21.4	41.8	666.2	46.7
			49.5	78.2	21.13	6.65	43.5	26.1	43.7	67.4	48.3
	10	FSRW FRCN+ft-full	33.4	61.4	30.09	91.57	14.36	64.21	30.4	53.4	28.6
			36.4	62.8	3.37	44.95	89.56	86.36	33.2	51.6	35.4
			47.6	74.6	1.03	65.49	64.19	21.02	42.5	67.9	44.7
			51.2	77.1	40.98	35.79	18.97	85.65	45.2	71.5	48.7
	15	FRCN+ft-full TFA w/fc	34.6	58.4	7.69	67.34	1.12	8.03	34.5	54.6	34.6
			48.5	64.2	45.19	78.74	79.79	51.39	41.6	68.9	47.6
			49.6	73.5	9.35	66.57	60.02	56.49	44.7	72.5	49.8
Split 2	1	FSRW	68.40	50.86	74.02	89.28	92.61	77.69	29.83	74.99	80.29
			36.33	41.29	94.91	26.98	32.09	16.37	37.50	38.14	78.07
			6.53	42.44	7.90	29.34	74.23	63.36	47.32	40.81	8.13
			71.55	48.29	6.30	88.85	15.55	88.45	86.07	74.72	17.83
	3	FSRWFRCN+ft-full	98.33	63.36	76.23	94.20	80.75	35.45	77.17	87.28	75.40
			95.83	77.31	28.75	21.51	23.65	4.80	73.60	99.00	73.48
			53.30	37.80	5.48	24.98	67.47	50.82	5.67	17.41	51.22
			4.23	63.71	6.37	21.37	70.55	8.84	90.17	47.71	70.55
	5	FSRW FRCN+ft-full	70.45	94.31	44.50	63.52	67.00	48.66	71.92	68.73	98.36
			76.78	83.49	90.09	36.31	79.07	73.35	54.64	54.79	9.36
			87.42	71.70	87.11	76.48	66.86	42.74	16.88	53.99	25.21
			97.98	90.51	62.83	36.52	38.50	78.10	28.77	89.60	3.11
	10	FSRW	82.06	40.71	70.36	96.43	32.87	14.28	25.45	3.04	67.73

		FRCN+ft-full	75.53	57.28	88.75	73.29	27.94	39.91	88.15	90.84	64.30
			60.59	84.96	39.50	35.29	56.97	71.82	59.43	77.68	75.81
			25.98	98.87	32.70	36.53	91.73	75.71	7.57	18.07	67.24
	15	FRCN+ft-full TFA w/fc	43.74	3.42	10.06	17.78	92.30	69.42	78.09	8.53	92.76
			80.38	32.25	33.08	67.12	72.15	38.41	65.41	51.63	27.59
			67.68	82.44	86.75	12.10	32.51	99.77	73.87	54.66	55.62
split 3	1	FSRW	84.91	98.03	33.81	6.40	41.56	43.73	5.75	78.08	7.10
			57.80	30.72	60.46	29.37	20.72	55.70	81.29	54.45	34.58
			71.34	11.17	78.12	29.38	65.96	51.52	68.94	73.14	63.25
			26.79	42.21	80.26	17.43	40.11	17.28	77.58	40.41	14.28
	3	FSRW FRCN+ft-full	49.35	64.03	83.86	19.22	99.22	98.38	72.82	29.79	11.27
			95.95	78.35	14.00	79.07	93.03	70.43	49.45	11.93	3.83
			28.22	85.36	18.92	6.42	96.13	82.51	31.76	71.24	24.26
			39.00	15.51	74.25	88.05	87.33	83.19	80.37	22.17	0.53
	5	FSRW FRCN+ft-full	48.94	40.09	42.60	26.55	48.86	12.96	62.71	31.90	94.12
			6.56	68.38	54.90	92.22	16.05	66.26	24.51	72.60	69.05
			67.69	45.04	27.19	58.80	76.33	32.05	45.79	69.48	20.21
			29.28	91.64	67.48	12.20	79.81	94.92	64.79	29.67	16.78
	10	FSRW FRCN+ft-full	91.77	48.64	80.06	54.65	23.19	17.05	77.56	5.18	50.22
			22.22	80.42	54.10	74.70	80.11	0.53	25.48	45.87	63.04
			82.50	58.39	92.48	56.83	29.73	12.54	35.96	74.93	78.11
			84.75	81.60	12.70	15.35	39.88	44.15	0.27	94.66	98.74
	15	FRCN+ft-full TFA w/fc	88.91	46.82	7.70	78.68	26.48	92.46	68.26	36.29	17.21
			83.80	32.36	97.31	43.20	33.03	61.87	50.15	13.88	99.18
			5.26	65.21	69.36	45.40	64.17	72.94	91.16	4.38	5.64

Table 2: Object detection on PASCAL VOC

Table 2 represents the complete revised benchmark on PASCAL VOC dataset with 93% confidence interval. In the tables, we show the average AP, AP^{50}, and AP^{75} for base classes, novel classes only, and overall classes.

The 93% confidence is calculated using equation 2:

$$93\%\ of\ class\ interval = 1.94 * \frac{s}{\sqrt{n}} \qquad (2)$$

Here, 1.94 is the z-value and standard deviation is denoted by s and n is number of iterations.

#shots	Method	Base class			Novel class			Overall					
		bAP	**bAP^{50}**	**bAP^{75}**	**nAP**	**nAP^{50}**	**nAP^{75}**	**AP**	**AP^{50}**	**AP^{75}**	**AP**	**APm**	**APl**
1	FRCN+ft-full	12.19	6.07	85.60	38.66	98.34	42.56	42.27	31.25	64.60	77.45	71.42	22.25
	TFA w/fc	68.06	7.84	92.01	99.21	83.78	8.90	61.09	89.83	75.33	60.75	40.09	13.23
	TFA w/cos	6.22	42.38	48.99	37.64	71.61	78.79	90.89	63.27	43.78	33.68	84.29	92.27
2	FRCN+ft-full	91.59	84.39	39.74	94.77	24.86	64.92	36.01	78.21	42.93	8.46	4.97	92.17
	TFA w/fc	99.76	72.98	52.38	98.77	14.17	64.31	89.61	82.53	25.12	65.40	42.04	75.86
	TFA w/cos	26.91	71.41	31.45	95.14	98.00	99.81	90.57	9.83	47.11	91.92	3.22	20.91
3	FRCN+ft-full	11.85	72.70	16.48	38.44	40.88	84.80	73.94	67.16	41.18	27.34	62.34	19.03
	TFA w/fc	51.92	38.38	41.27	40.46	7.58	90.71	37.06	84.57	23.75	47.84	90.38	24.90
	TFA w/cos	66.73	41.45	43.01	11.75	35.19	38.38	68.41	83.33	95.87	61.29	79.89	49.42
5	FRCN+ft-full	59.92	76.59	44.31	28.80	52.10	9.83	89.84	61.24	19.94	71.51	13.98	7.18
	TFA w/fc	1.72	66.44	75.55	84.63	52.11	94.89	13.03	15.66	25.57	49.08	91.29	60.83
	TFA w/cos	34.64	30.51	64.41	5.59	28.96	95.36	93.89	17.21	49.26	89.63	5.92	72.84
10	FRCN+ft-full	68.65	23.30	81.02	9.17	26.02	75.24	40.82	42.74	12.65	18.28	61.00	7.05
	TFA w/fc	29.35	17.11	73.71	91.26	92.35	90.61	71.16	6.34	97.88	77.53	74.39	53.11
	TFA w/cos	15.30	14.96	30.72	28.00	88.76	76.89	93.80	84.39	88.99	32.40	82.44	97.57
30	FRCN+ft-full	63.66	82.65	78.77	5.65	74.46	70.82	28.67	21.88	20.49	64.08	96.16	41.70
	TFA w/fc	30.17	19.16	79.85	92.87	31.76	82.48	90.79	5.68	22.95	29.34	31.49	43.11
	TFA w/cos	51.69	53.85	36.79	36.85	78.48	61.08	91.02	96.86	46.64	36.96	71.00	77.90

Table 3: Object detection on COCO

The model was evaluated for different shots and consistently outperformed existing model by 4 to 16 points in the novel class.Table 3 represents the complete revised benchmark on COCO dataset with 93% confidence interval. The n value is 15 for the COCO dataset. We have presented the average precisions (AP) on 50 and 75 interval i.e. AP^{50} and AP^{75} for base, novel and overall classes.

Conclusion

We described a loss function-based technique for two-stage few-shot object recognition and rating, which can rationally and efficiently execute both tasks. On a variety of well-known benchmarks for few-shot object recognition, we exhibited the advantages and accuracy of our technique, and we greatly improved the state-of-the-art. Our models attained a recognition rate of 32.3 percent on the PASCAL VOC dataset and 39.5 percent on the COCO dataset. We also demonstrated that our few-shot model can generate impressive outcomes on unique objects discovered by our few-shot detector and we confirm our findings by comparing it to ground truth provided by the existing approaches.

References

Berkeley, U. C., Girshick, R., & Berkeley, U. C. (2017). Contextual Action Recognition with R * CNN. *Iccv*, 1080–1088.

Chabot, F., Pham, Q.-C., & Chaouch, M. (2019). *LapNet : Automatic Balanced Loss and Optimal Assignment for Real-Time Dense Object Detection*. http://arxiv.org/abs/1911.01149

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *34th International Conference on Machine Learning, ICML 2017*, *3*, 1856–1868.

Gidaris, S., & Komodakis, N. (2018). Dynamic Few-Shot Visual Learning Without Forgetting. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4367–4375. https://doi.org/10.1109/CVPR.2018.00459

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/CVPR.2014.81

Jiang, W., Huang, K., Geng, J., & Deng, X. (2021). Multi-Scale Metric Learning for Few-Shot Learning. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(3), 1091–1102. https://doi.org/10.1109/TCSVT.2020.2995754

Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., & Darrell, T. (2019). Few-shot object detection via feature reweighting. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*(Iccv), 8419–8428. https://doi.org/10.1109/ICCV.2019.00851

Karlinsky, L., Shtok, J., Harary, S., Schwartz, E., Aides, A., Feris, R., Giryes, R., & Bronstein, A. M. (2019). Repmet: Representative-based metric learning for classification and few-shot object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 5192–5201. https://doi.org/10.1109/CVPR.2019.00534

Kolmogorov, V., & Rol, M. (n.d.). *Efficient Optimization for Rank-based Loss Functions Supplementary Material*. *1*(c), 3693–3701. http://openaccess.thecvf.com/content_cvpr_2018/papers/Mohapatra_Efficient_Optimization_for_CVPR_2018_paper.pdf

Li, A., Huang, W., Lan, X., Feng, J., Li, Z., & Wang, L. (2020). Boosting Few-Shot Learning with Adaptive Margin Loss. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 12573–12581.

https://doi.org/10.1109/CVPR42600.2020.01259

Li, A., Luo, T., Xiang, T., Huang, W., & Wang, L. (2019). Few-shot learning with global class representations. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 9714–9723. https://doi.org/10.1109/ICCV.2019.00981

Li, X., Yu, L., Fu, C. W., Fang, M., & Heng, P. A. (2020). Revisiting metric learning for few-shot image classification. *Neurocomputing*, *406*, 49–58. https://doi.org/10.1016/j.neucom.2020.04.040

Liu, B., Kang, H., Li, H., Vasconcelos, N., & Hua, G. (2020). Few-shot open-set recognition using meta-learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8795–8804. https://doi.org/10.1109/CVPR42600.2020.00882

Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., & Fan, D.-P. (2021). *Simultaneously Localize, Segment and Rank the Camouflaged Objects*. http://arxiv.org/abs/2103.04011

Oksuz, K., Cam, B. C., Akbas, E., & Kalkan, S. (2018). Localization recall precision (LRP): A new performance metric for object detection. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11211 LNCS*, 521–537. https://doi.org/10.1007/978-3-030-01234-2_31

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Samuelson, L. K., & Smith, L. B. (2005). They call it like they see it: Spontaneous naming and attention to shape. *Developmental Science*, *8*(2), 182–198. https://doi.org/10.1111/j.1467-7687.2005.00405.x

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*(1), 13–19. https://doi.org/10.1111/1467-9280.00403

Tan, Z., Nie, X., Qian, Q., Li, N., & Li, H. (2019). Learning to rank proposals for object detection. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 8272–8280. https://doi.org/10.1109/ICCV.2019.00836

Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., & Wierstra, D. (2016). Matching networks for one shot learning. *Advances in Neural Information Processing Systems*, 3637–3645.

Wang, T., Zhang, X., Yuan, L., & Feng, J. (2019). Few-shot adaptive faster R-CNN. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 7166–7175. https://doi.org/10.1109/CVPR.2019.00734

Wang, X., Huang, T. E., Darrell, T., Gonzalez, J. E., & Yu, F. (2020). Frustratingly simple few-shot object detection. *37th International Conference on Machine Learning, ICML 2020*, *PartF16814*, 9861–9870.

Wang, Y. X., Ramanan, D., & Hebert, M. (2019). Meta-learning to detect rare objects.

Proceedings of the IEEE International Conference on Computer Vision, *2019-Octob*, 9924–9933. https://doi.org/10.1109/ICCV.2019.01002

Xu, Z., Li, B., Yuan, Y., Research, M., & Dang, A. (2020). Beta R-CNN: Looking into Pedestrian Detection from Another Perspective. *Advances in Neural Information Processing Systems*, *33*(NeurIPS), 19953–19963.

Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X., & Lin, L. (2019). Meta R-CNN: Towards general solver for instance-level low-shot learning. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 9576–9585. https://doi.org/10.1109/ICCV.2019.00967