華東師範大學 EAST CHINA NORMAL UNIVERSITY

腾讯优图

Mutually Reinforcing Structure with Proposal Contrastive Consistency for Few-Shot Object Detection

Tianxue Ma[1], Mingwei Bi[2], Jian Zhang[2], Wang Yuan[1], Zhizhong Zhang[1], Yuan Xie[1], Shouhong Ding[2], Lizhuang Ma[1]

[1] East China Normal University, Shanghai, China, [2] Tencent Youtu Lab, Shanghai, China

Introduction

➢ Background

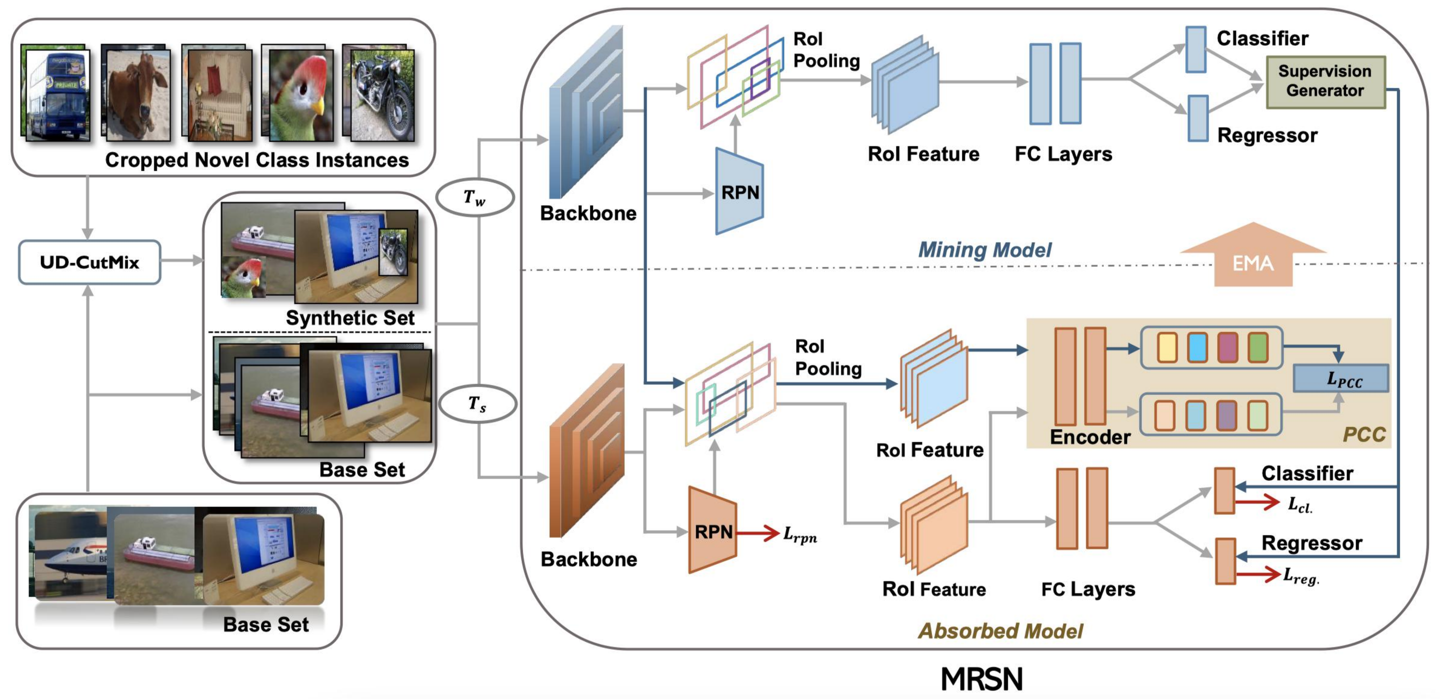
- The application of deep convolutional neural network has accelerated the development of object detection. However, the remarkable performance of object detection depends on abundant annotated data. Collecting annotated data is time-consuming and labor-intensive. As opposed to that, a few examples are sufficient for humans to learn a new concept. To bridge this gap, we focus on few-shot object detection (FSOD), which aims to adapt the model from the base class to the novel class based on a few labeled novel classes examples.
- Based upon the base set, almost all methods for solving FSOD simply utilize the labeled base class objects in the base set as supervision information to train a detector. In the process of using base set, past methods overlook an important phenomenon: material neglect, in which unlabeled novel-class instances are explicitly learned as background.
- To address the above issue, we propose a Mutually Reinforcing Structure Network (MRSN) to make rational use of unlabeled novel class instances in the base set.

➢ Our Contributions

- We propose a Mutually Reinforcing Structure Network (MRSN) to make rational use of unlabeled novel class instances in the base set. MRSN consists of a mining model which unearths unlabeled novel-class instances and an absorbed model which learns variable knowledge.
- Then, we design a Proposal Contrastive Consistency (PCC) module in the absorbed model to fully exploit class characteristics and avoid bias from unearthed labels. Furthermore,we propose a simple and effective data synthesis method undirectional-CutMix (UD-CutMix) to improve the robustness of model mining novel class instances.
- Extensive experiments and visualizations reveal the significance of making rational use of unlabeled novel class instances in base set, and demonstrate the effectiveness of our proposed method against the state-of-the-art competitors.

Method

➢ Network structure



MRSN

- Our MRSN is a dual model construction, including a mining model and an absorbed model , where mining model is used to mine unlabeled novel-class instances and give them pseudo labels including corresponding categories and locations, and absorbed model is used to learn the mined instances.
- UD-CutMix combines the cropped novel classes instances and the selected image in base set. UD-CutMix adopts the detector to select the base image, which prediction doesn't contain any novel objects, to be mixed. Specifically, we first crop a novel-class instance from a novel set image, then scale and paste it to a selected base image. By repeating this operation, we can construct a new synthetic set .
- We propose a contrastive learning method Proposal Contrastive Consistency (PCC) suitable for our MRSN in FSOD. We introduce a PCC branch to the primary RoI head, parallel to the classification and regression branches. According to RoIs sampled by absorbed mode , RoIPooling is performed on the feature maps obtained by the feature extractors of mining model and absorbed model to obtain the features of the proposals. We implement contrastive learning by taking the features from the same proposal and different models as positive sample pairs and others as negative sample pairs.

Experiments

➢ Performance on the PASCAL VOC novel classes

Method / Shot	Novel Set 1					Novel Set 2					Novel Set 3				
	1-shot	2-shot	3-shot	5-shot	10-shot	1-shot	2-shot	3-shot	5-shot	10-shot	1-shot	2-shot	3-shot	5-shot	10-shot
MetaDet [39]	17.1	19.1	28.9	35.0	48.8	18.2	20.6	25.9	30.6	41.5	20.1	22.3	27.9	41.9	42.9
RepMet [15]	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
CRDR [21]	40.7	45.1	46.5	57.4	62.4	27.3	31.4	40.8	42.7	46.3	31.2	36.4	43.7	50.1	55.6
FRCN-ft [39]	13.8	19.6	32.8	41.5	45.6	7.9	15.3	26.2	31.6	39.1	9.8	11.3	19.1	35.0	45.1
Meta R-CNN [44]	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
TFA w/ fc [38]	36.8	29.1	43.6	55.7	57.0	18.2	29.0	33.4	35.5	39.0	27.7	33.6	42.5	48.7	50.2
MPSR [42]	41.7	-	51.4	55.2	61.8	24.4	-	39.2	39.9	47.8	**35.6**	-	42.3	48.0	49.7
FSCE [33]	32.9	44.0	46.8	52.9	59.7	23.7	30.6	38.4	43.0	48.5	22.6	33.4	39.5	47.3	54.0
CME [20]	41.5	47.5	50.4	58.2	60.9	27.2	30.2	41.4	42.5	46.8	34.3	39.6	45.1	48.3	51.5
DCNet [13]	33.9	37.4	43.7	51.1	59.6	23.2	24.8	30.6	36.7	46.6	32.3	34.9	39.7	42.6	50.7
UP [41]	43.8	47.8	50.3	55.4	61.7	31.2	30.5	41.2	42.2	48.3	35.5	**39.7**	43.9	50.6	53.5
Ours	**47.6**	**48.6**	**57.8**	**61.9**	**62.6**	**31.2**	**38.3**	**46.7**	**47.1**	**50.6**	35.5	30.9	**45.6**	**54.4**	**57.4**

➢ Performance on MS-COCO novel classes

Method / Shot	10-shot		30-shot	
	AP	AP75	AP	AP75
MetaDet 2019 [39]	7.1	6.1	11.3	8.1
Meta R-CNN 2019 [44]	8.7	6.6	12.4	10.8
TFA w/fc 2020 [38]	9.1	8.8	12.1	12.0
MPSR 2020 [42]	9.8	9.7	14.1	14.2
Viewpoint [43]	12.5	9.8	14.7	12.2
FSCE 2021 [33]	11.1	9.8	15.3	14.2
CRDR 2021 [21]	11.3	-	15.1	-
UP 2021 [41]	11.0	10.7	15.6	15.7
Ours	**15.7**	**14.8**	**17.5**	**17.9**

Conclusion

➢ In this paper, we discover that previous FSOD solving methods overlook material neglect. Towards solving this problem, a Mutually Reinforcing Structure Network (MRSN) is introduced to mine and absorb unlabeled novel instances. We design a Proposal Contrastive Consistency (PCC) module in MRSN to help learn more detailed features and resist the influence of noise labels. Meanwhile, we design a data synthesis method undirectional-CutMix (UD-CutMix) that combines the novel class instances and the images in the base set. Our method can effectively solve the problem of material neglect in FSOD. Experiments show that our method outperforms the previous methods by a large margin.