dongguk UNIVERSITY

Meta R-CNN : Towards General Solver for Instance-level Few-shot Learning

2021.11.4

Ershang Tian

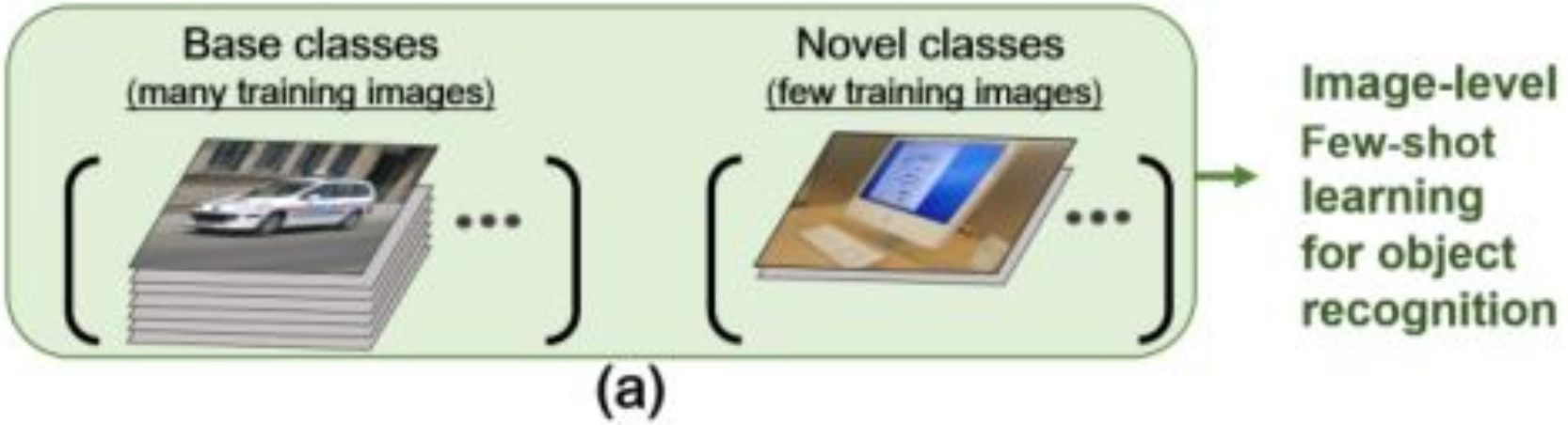
Introduction

- Obfuscated by a complex background and multiple objects in one image, they are hard to promote the research of few-shot object detection/segmentation.

- Extends Faster /Mask R-CNN by proposing meta-learning over RoI (Region-of-Interest) features instead of a full image feature.

Introduction

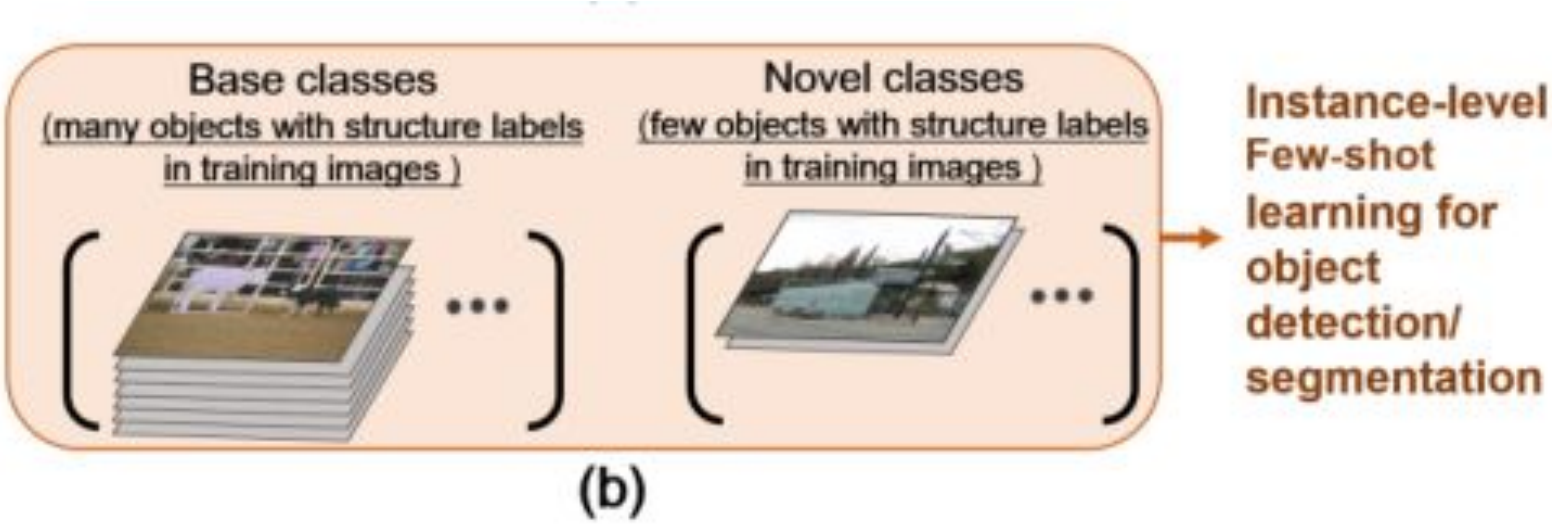
- Provided with very few labeled data (1～10 shots) in novel classes, low-shot learners are trained to recognize the data-starveclass objects with the aid of base classes with sufficient labeled data



Introduction

- In terms of instance-level learning tasks, e.g., object detection/segmentation, prior works in low-shot learning contexts remain rarely explored

- It is more difficult to detect/segment objects with multi-structure tags than to annotate images



동국대학교
dongguk university

Introduction

- **A novel meta-learning paradigm based on the RoI (Region-of-Interest) features produced by Faster/Mask R-CNN.**
  - Blended undiscovered objects could be “pre-processed” via the RoI features produced by the first-stage inference in Faster /Mask R-CNNs.

- **PRN（Region Proposal Network）:**
  - PRN is fully-convoluted
  - shares the main backbone’s parameters with Faster /Mask R-CNN.

동국대학교 dongguk university

Tasks and Motivation

- **low-shot visual object recognition**

- **In few-shot object recognition, a learner *h*(;*θ*) receives training data from *base classes* C_{base} and *novel classes* C_{novel}. So the data can be divided into two groups:**

- $D_{\text{base}} = \{(\mathbf{x}_i^{\text{base}}, y_i^{\text{base}})\}_{i=1}^{n_1} \sim P_{\text{base}}$

- $D_{\text{novel}} = \{(\mathbf{x}_i \quad , y_i \quad)\}_{i=1}^{2} \sim P_{\text{novel novel }}\}n$ novel

동국대학교
dongguk university

Tasks and Motivation

- **low-shot visual object recognition**
  - $h(;\theta)$ aims to classify test samples drawn from P_{novel}.
  - $h(;\theta)$ with small dataset D_{novel} to identify C_{novel} suffers model overfitting, whereas training $h(;\theta)$ with $D_{\text{base}} \cup D_{\text{novel}}$ still fails, due to the extreme data quantity imbalance between Dbase and Dnovel ($n2 << n1$).

동국대학교
dongguk university

Tasks and Motivation

-
- Few-shot object detection / segmentation
- **In the two-stage detection model, ROL features pass through the region proposal network (RPN). It is input into the predictor head of the second stage to achieve RoI-based object classification, location positioning and contour segmentation. In view of this, it is best to remodel the head of the R-CNN predictor as h(zi,j,Dmeta;θ) in order to perform the process for each region of interest (RoI) feature ˆzi,j. behind the object zi,j. Classification, positioning and segmentation.**

동국대학교
dongguk university

Meta R-CNN

- Meta R-CNN is conceptually simple: its pipeline consists of 1). Faster/ Mask R-CNN; 2). Predictor-head Remodeling Network (PRN).

- Faster/ Mask R-CNN produces object proposals $\{\hat{\mathbf{z}}_{i,j}\}_{j=1}^{n_i}$ by their region proposal networks (RPN).

- Each ˆzi,j combines with class-attentive vectors inferred by our PRN, which plays the role of h(, Dmeta;Θ) to detect or segment the novel-class objects.

동국대학교
dongguk university

Meta R-CNN

- Predictor-head Remodeling Network (PRN)

- Meta R-CNN is conceptually simple: its pipeline consists of 1). Faster/ Mask R-CNN; 2). Predictor-head Remodeling Network (PRN).

- Faster/ Mask R-CNN produces object proposals $\{\hat{\mathbf{z}}_{i,j}\}_{j=1}^{n_i}$ by their region proposal networks (RPN).

- Each ˆzi,j combines with class-attentive vectors inferred by our PRN, which plays the role of h(, $\boldsymbol{D}_{\text{meta}}$;Θ) to detect or segment the novel-class objects.

동국대학교
dongguk university

Implementation

- **Meta R-CNN is trained under a meta-learning paradigm.**

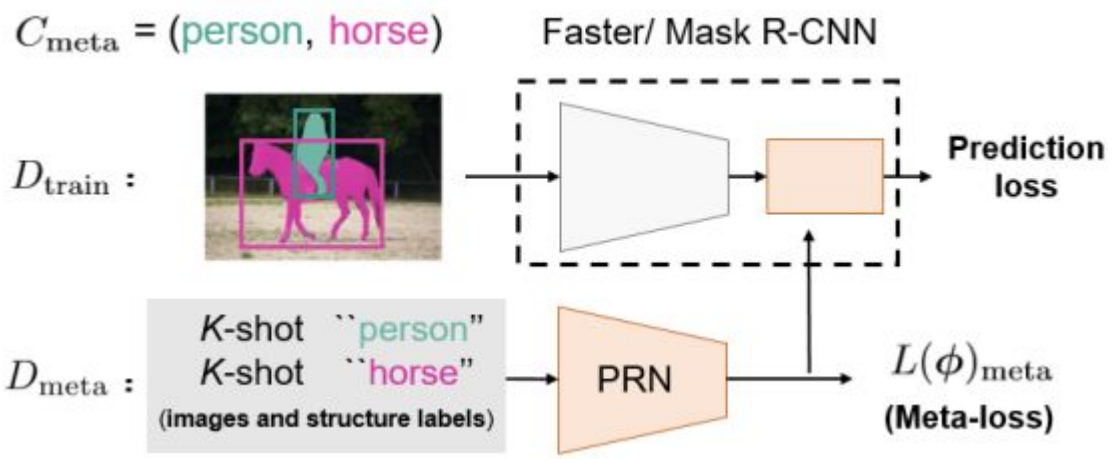
- Mini-batch construction. Simulating the meta-learning paradigm we have discussed, a training mini-batch in Meta R-CNN is comprised of m classes Cmeta ~ Cbase ∪Cnovel, a K-shot m-class meta-set Dmeta and m-class training set Dtrain

- R-CNN module receives an image input x that contains objects in m classes, a mini-batch consists of x (Dtrain) and mK resized images with their structure label masks.

동국대학교
dongguk university

Implementation

- **Meta R-CNN entails two inference processes based on Faster/Mask R-CNN module and PRN.**



동국대학교
dongguk university

Experiments

Table 3. AP and mAP on VOC2007 test set for novel classes and base classes of the first base/novel split. We evaluate the performance for 3/10-shot novel-class examples with FRCN under ResNet-101. RED/BLUE indicate the SOTA/the second best. (Best viewd in color)

		Novel classes						Base classes														mAP
Shot	Baselines	bird	bus	cow	mbike	sofa	mean	aero	bike	boat	bottle	car	cat	chair	table	dog	horse	person	plant	sheep	train	tv	mean	
3	YOLO-Few-shot [21]	26.1	19.1	40.7	20.4	**27.1**	26.7	**73.6**	**73.1**	56.7	41.6	**76.1**	78.7	42.6	**66.8**	72.0	77.7	68.5	**42.0**	57.1	**74.7**	70.7	**64.8**	55.2
	FRCN+joint	13.7	0.4	6.4	0.8	0.2	4.3	**75.9**	**80.0**	**65.9**	**61.3**	**85.5**	**86.1**	**54.1**	**68.4**	**83.3**	**79.1**	**78.8**	**43.7**	**72.8**	**80.8**	**74.7**	**72.7**	55.6
	FRCN+ft	**31.1**	24.9	**51.7**	23.5	13.6	29.0	65.4	56.4	46.5	41.5	73.3	84.0	40.2	55.9	72.1	75.6	74.8	32.7	60.4	71.2	**71.2**	61.4	53.3
	FRCN+ft-full	29.1	**34.1**	**55.9**	**28.6**	**16.1**	**32.8**	67.4	62.0	54.3	48.5	74.0	**85.8**	42.2	58.1	72.0	**77.8**	75.8	32.3	**61.0**	73.7	68.6	63.6	**55.9**
	Meta R-CNN (ours)	**30.1**	**44.6**	50.8	**38.8**	10.7	**35.0**	67.6	70.5	**59.8**	**50.0**	75.7	81.4	**44.9**	57.7	**76.3**	74.9	**76.9**	34.7	58.7	**74.7**	67.8	**64.8**	**57.3**
10	YOLO-Few-shot [21]	30.0	**62.7**	43.2	**60.6**	39.6	**47.2**	65.3	73.5	54.7	39.5	75.7	81.1	35.3	62.5	72.8	78.8	68.6	**41.5**	59.2	**76.2**	69.2	63.6	**59.5**
	FRCN+joint	14.6	20.3	19.2	24.3	2.2	16.1	**78.1**	**80.0**	**65.9**	**64.1**	**86.0**	**87.1**	**56.9**	**69.7**	**84.1**	**80.0**	**78.4**	**44.8**	**74.6**	**82.7**	**74.1**	**73.8**	59.4
	FRCN+ft	31.3	36.5	**54.1**	26.5	36.2	36.9	**68.4**	**75.2**	59.2	**54.8**	74.1	80.8	42.8	56.0	68.9	77.8	75.5	34.7	**66.1**	71.2	66.2	64.8	57.8
	FRCN+ft-full	**40.1**	47.8	45.5	47.5	**47.0**	45.6	65.7	69.2	52.6	46.5	74.6	73.6	40.7	55.0	69.3	73.5	73.2	33.8	56.5	69.8	65.1	61.3	57.4
	Meta R-CNN (ours)	**52.5**	**55.9**	**52.7**	**54.6**	**41.6**	**51.5**	68.1	73.9	**59.8**	54.2	**80.1**	**82.9**	**48.8**	**62.8**	**80.1**	**81.4**	**77.2**	37.2	65.7	75.8	**70.6**	**67.9**	**63.8**

동국대학교
dongguk university

Experiments

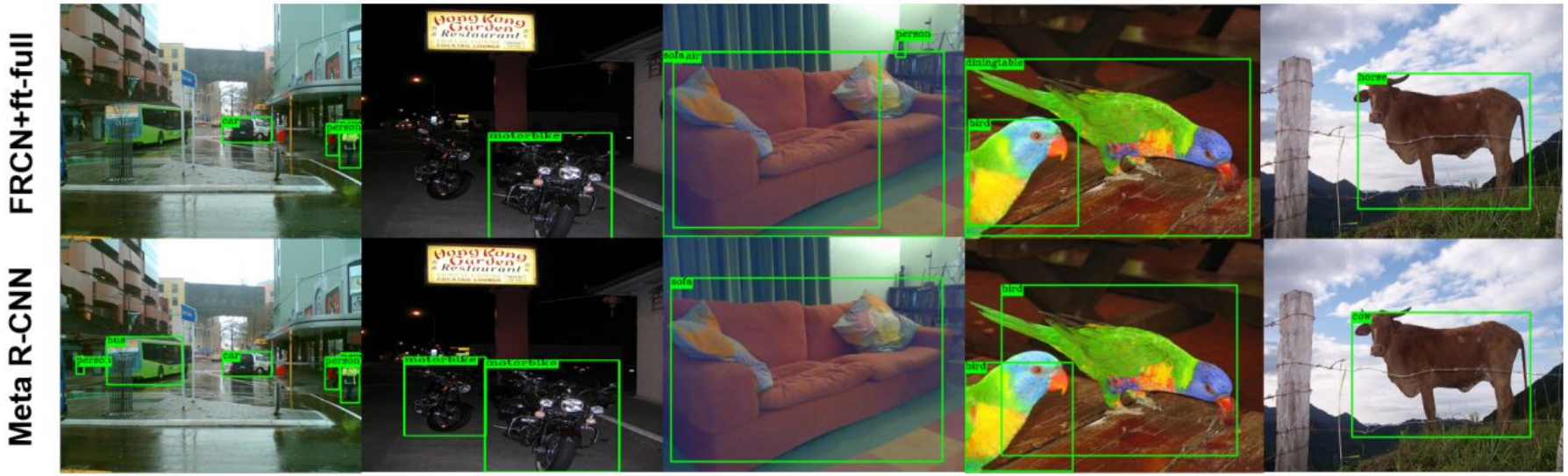


Figure 4. The visualization of novel-class objects detected by FRCN+ft-full and Meta R-CNN. Compared with Meta R-CNN, FRCN+ft-full is inferior: bboxes in the first two columns are missed; in the middle column is duplicate and the classes are wrong in the last two columns.

동국대학교
dongguk university

Discussion and Future Work

- Few-shot object detection/ segmentation are very valuable
- Meta R-CNN overcomes the shared weakness of existing
- It endows traditional Faster/ Mask R-CNN with the generalization capability in front of few-shot objects in novel classes.

THANKS

Q&A

Thank you!