Code!

CRAFT Objects from Images

Bin Yang[1], Junjie Yan[2], Zhen Lei[1], Stan Z. Li[1]
[1]NLPR, CASIA, [2]Tsinghua University
http://byangderek.github.io/projects/craft.html

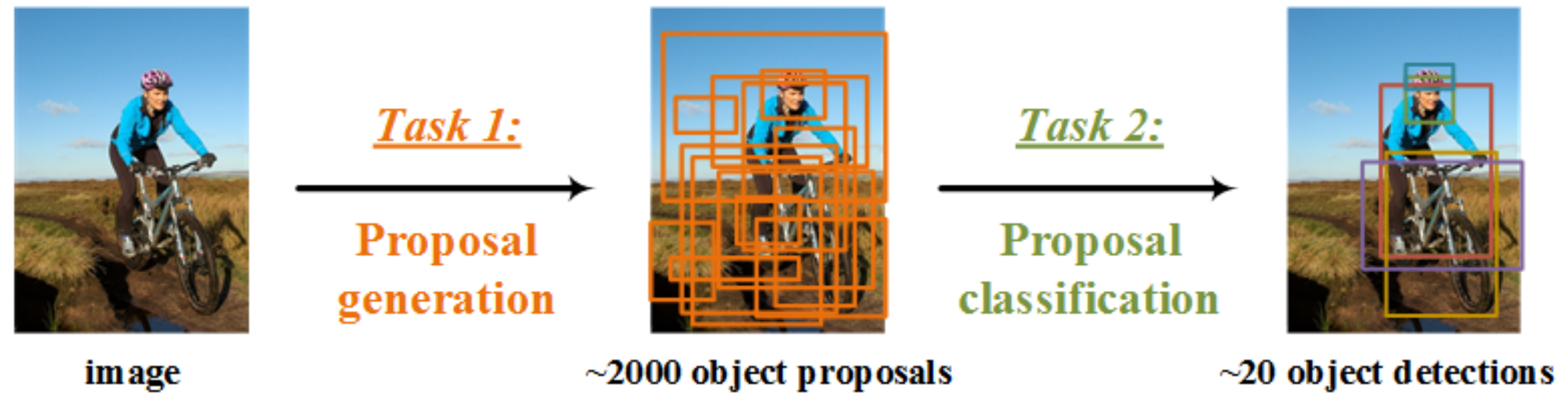
Code!

IEEE 2016 Conference on
Computer Vision and Pattern
Recognition
CVPR 2016

Adopting the *two-step* detection framework? Why don't we take *more baby steps*?

Motivation

The two-step detection framework



Gap between *ideal* and *reality*

➢ **Task 1: Proposal generation**

Ideal:

- Output only object proposals.

Reality:

Method	#Regions	Background regions	Recall@ 0.5IoU	Recall@ >0.8IoU	Recall@ hard_object
Selective Search	☹	☹	😐	😃	😃
RPN	😐	😐	😃	☹	☹

➢ **Task 2: Object classification**

Ideal:

- Classify proposals into N object categories of interest.

Reality:

- A majority of samples are background ($num\_classes$ becomes N+1);
- Samples of N different categories may vary a lot;
- With cross-entropy objective, CNN learns biased representation, and it is hard to capture fine-grained variance of each category.

Wrong detections of Fast R-CNN:

'potted_plant', 0.6

'tv_monitor', 0.8

'boat', 0.7

'dog', 0.5

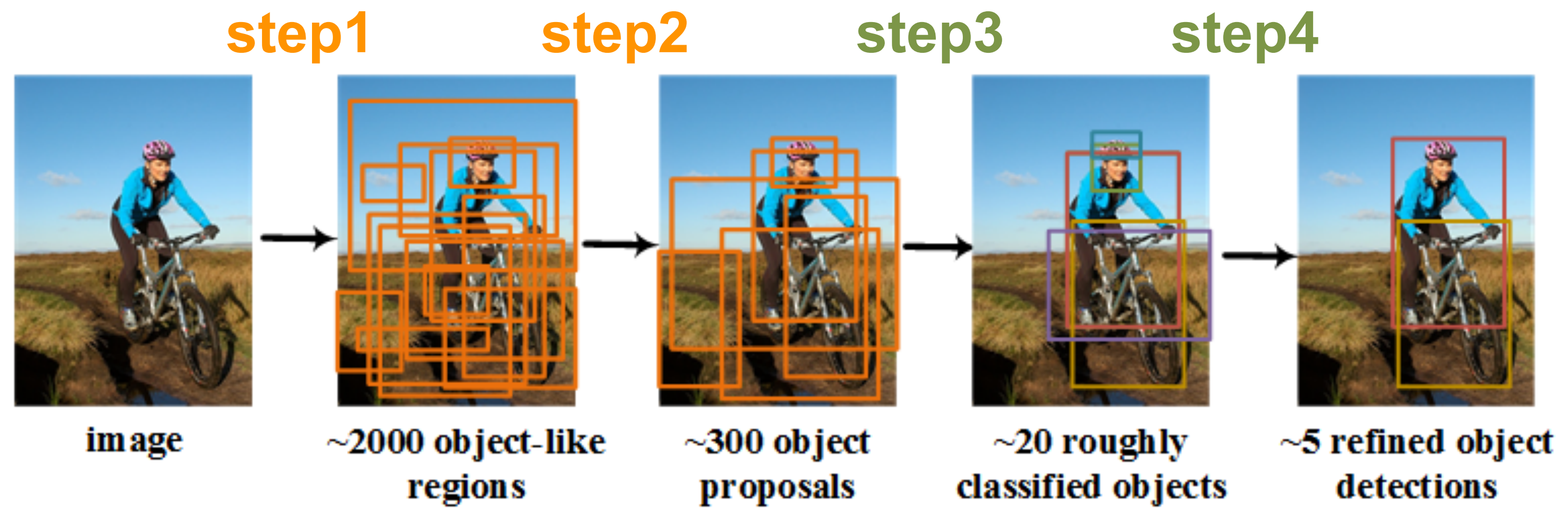
Solution

➢ Using *'divide and conquer' philosophy* to further decompose and better solve each of the two tasks;

➢ Each task is solved with a carefully designed neural network cascade.

Approach

➢ CRAFT (Cascade Rpn And FasT-rcnn)


step1
step2
step3
step4
image
~2000 object-like regions
~300 object proposals
~20 roughly classified objects
~5 refined object detections

Definition:

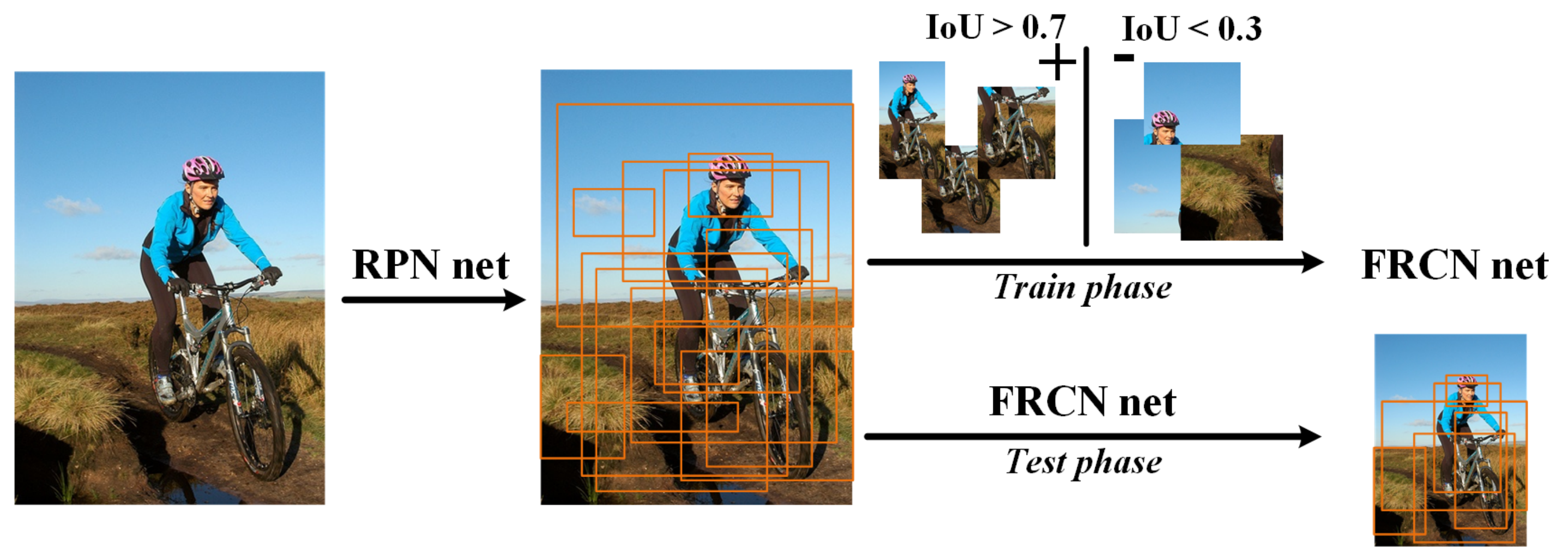
step1: standard RPN

step2: binary Fast R-CNN

step3: standard Fast R-CNN

step4: Fast R-CNN with N binary classifiers

➢ **Cascade proposal generation (step1 + step2)**

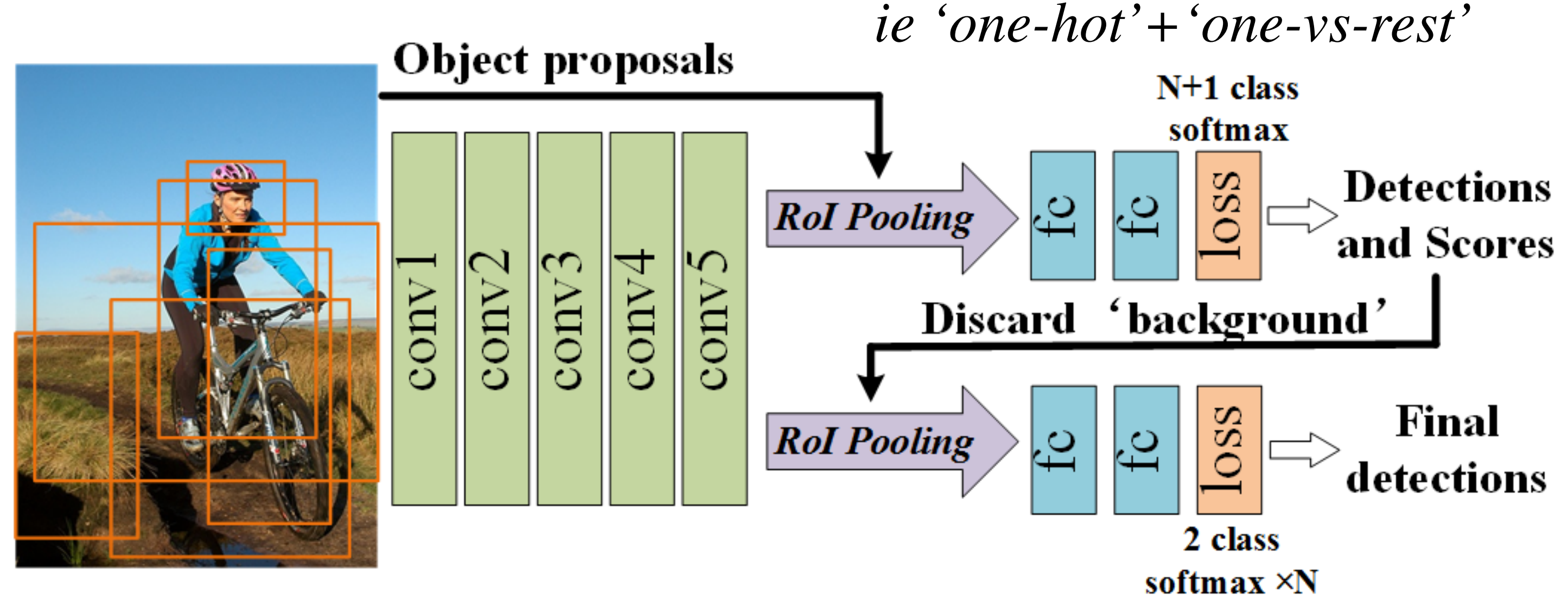


Advantage:

- eliminate *difficult* background regions;
- improve localization;
- combine proposals from multiple sources;
- 20% absolute recall gain at 0.8IoU with 5% proposals, 1% absolute mAP gain.

➢ **Cascade object classification (step3 + step4)**

ie 'one-hot'+'one-vs-rest'



Advantage:

- share full-image features;
- capture both *inter-* and *intra-* category variances;
- eliminate false positives between ambiguous categories;
- 3% absolute mAP gain on VOC07.

Results

➢ **Object proposal on VOC07 test**

- **Recall analysis on difficult categories**

Method	#Boxes	Recall	bird	boat	bottle	chair	plant	tv
VGG_M	300	94.8	93.8	92.7	80.3	91.7	86.8	90.5
VGG_19	300	97.5	96.2	95.8	**92.3**	95.6	90.4	95.1
Cascade VGG_M	300	**97.9**	**97.3**	**96.9**	92.1	**96.2**	**94.5**	**98.3**

- **Recall analysis at various IoUs and the detection mAP**

Method	#Boxes	0.5	0.6	0.7	0.8	0.9	mAP
SS	2000	92.1	85.2	72.5	52.9	**26.6**	70.0
RPN	2000	**98.5**	**95.8**	84.1	40.7	4.1	-
RPN	300	96.3	92.5	78.8	37.9	3.9	71.6
Ours	300	97.9	95.5	**89.6**	**63.7**	13.0	72.2
Ours_S	**87**	96.8	94.1	87.8	62.4	12.9	**72.5**

➢ **Object detection on VOC07/12 test and ILSVRC val2**

Method	proposal	classifier	voc07	voc12	ilsvrc
FRCN	SS	FRCN	70.0	65.7	-
RPN_unshared	RPN	FRCN	71.6	65.5	45.4
RPN	RPN	FRCN	73.2	67.0	-
Ours	cascade	FRCN	72.5	-	47.0
Ours	cascade	cascade	**75.7**	**71.3**	**48.5**

➢ **ImageNet 2015 Object Detection from Video (VID) Competition**

Team	Task	Track	Detector	AP_val	Rank
CUvideo	VID	Provided data	Ours	**67.7**	**1**
			DeepID-net	65.8	

Discussion

➢ CRAFT enjoys other advances in object detection like ION, ResNet;

➢ The cascade structure used in proposal task plays the role of ***hard example mining*** for the following detection task;

➢ The cascade structure used in detection task points out a potential drawback of current ***loss function choice*** for fast r-cnn, and provides an alternative solution.