Cascade RPN: Delving into High-Quality Region Proposal Network with Adaptive Convolution

Thang Vu, Hyunjun Jang, Trung X. Pham, Chang D. Yoo
Korea Advanced Institute of Science and Technology

AIM Lab. Artificial Intelligence & Machine Learning Lab.

NeurIPS 2019

Abstract

- This paper considers an architecture referred to as Cascade Region Proposal Network (Cascade RPN).
- Cascade RPN uses single anchor per location and perform multistage refinement instead of using multiple anchors.
- Cascade RPN systematically maintains alignment through stages using proposed adaptive convolution
- Cascade RPN achieves state-of-the-art region proposal performance on COCO dataset.

Introduction

Problem description.

- Conventional RPN uses multiple scales and aspect ratios which are heuristically defines and requires tuning for good performance.

→ **Cascade RPN relies on a single anchor per location and performs multi-stage refinement.**

- Alignment between anchor box and feature is not well-preserved in existing multi-stage RPN.

→ **Cascade RPN relies on adaptive convolution.**

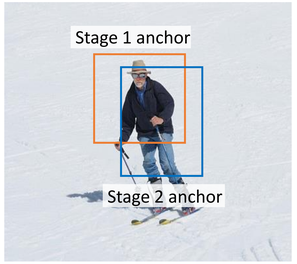
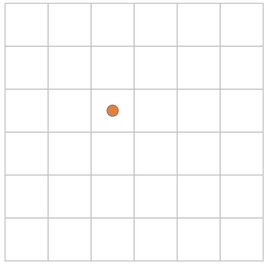


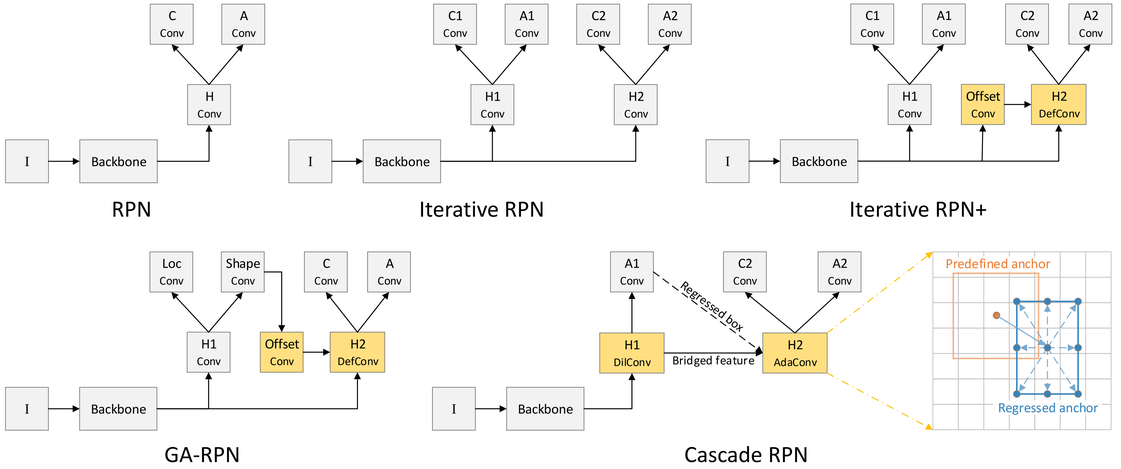
Image space



Feature space

Misalignment problem in iterative RPN

Cascade RPN

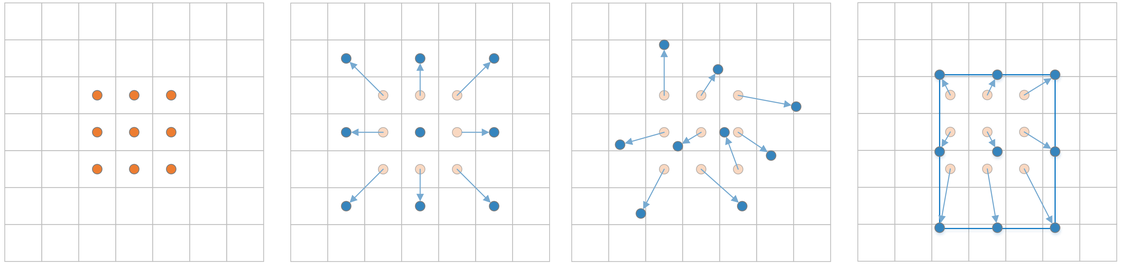


Architectures of Cascade RPN and others.

Adaptive convolution

- In standard convolution, feature weighted sum is performed on regular grid $\boldsymbol{y}[\boldsymbol{p}] = \sum_{\boldsymbol{r}\in\mathbb{R}} \boldsymbol{w}[\boldsymbol{r}] \cdot \boldsymbol{x}[\boldsymbol{p}+\boldsymbol{r}]$
- In adaptive convolution, the regular grid $\mathbb{R}$ is replaced by the offset grid $\mathbb{O}$: $\boldsymbol{y}[\boldsymbol{p}] = \sum_{\boldsymbol{o}\in\mathbb{O}} \boldsymbol{w}[\boldsymbol{o}] \cdot \boldsymbol{x}[\boldsymbol{p}+\boldsymbol{o}]$
- The offset o can be decoupled into center offset and shape offset: $\boldsymbol{o} = \boldsymbol{o}_{ctr} + \boldsymbol{o}_{shp}$

$\boldsymbol{o}_{ctr}$: position alignment $\boldsymbol{o}_{shp}$:semantic scope alignment



Standard Conv | Dilated Conv | Deformable Conv | Adaptive Conv

Sampling locations of Adaptive Convolution and others.

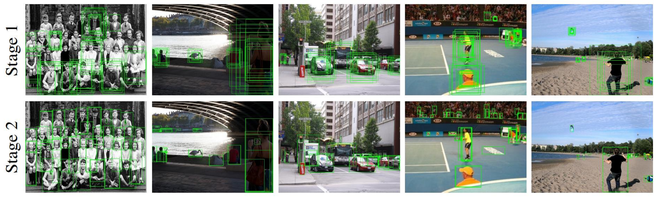
Experiments

Region proposal results on COCO 2017 val.

Method	Backbone	AR_{100}	AR_{300}	AR_{1000}	AR_S	AR_M	AR_L	Time (s)
SharpMask [30]	ResNet-50	36.4	-	48.2	-	-	-	0.76
GCN-NS [28]	VGG-16 (Sync BN)	31.6	-	60.7	-	-	-	0.10
AttractioNet [14]	VGG-16	53.3	-	66.2	31.5	62.2	77.7	4.00
ZIP [22]	BN-inception	53.9	-	67.0	31.9	63.0	78.5	1.13
RPN [34]	ResNet-50-FPN	44.6	52.9	58.3	29.5	51.7	61.4	**0.04**
Iterative RPN		48.5	55.4	58.8	32.1	56.9	65.4	0.05
Iterative RPN+		54.0	60.4	63.0	35.6	62.7	73.9	0.06
GA-RPN [37]		59.1	65.1	68.5	40.7	68.2	78.4	0.06
Cascade RPN		**61.1**	**67.6**	**71.7**	**42.1**	**69.3**	**82.8**	0.06

Ablation study of Cascade RPN.

Baseline	1 anchor	Cascade	Align.	AFAB	Stats.	IoU loss	AR_{100}	AR_{300}	AR_{1000}
✓							44.6	52.9	58.3
	✓						44.7	51.2	55.8
	✓	✓					48.2	54.4	58.0
	✓	✓	✓				57.4	63.7	67.8
	✓	✓	✓	✓			57.3	64.2	68.6
	✓	✓	✓	✓	✓		60.8	67.3	71.5
	✓	✓	✓	✓	✓	✓	**61.1**	**67.6**	**71.7**
Overall Improvement							**+16.5**	**+14.7**	**+13.4**



Qualitative results for stage 1 and stage 2 of Cascade RPN.

Detection results on COCO 2017 test-dev.

Method	Proposal method	# proposals	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Fast R-CNN	RPN	1000	37.0	**59.5**	39.9	21.1	39.4	47.0
	Cascade RPN		**40.1**	**59.5**	**43.7**	**22.8**	**42.4**	**50.9**
	RPN	300	36.6	58.6	39.5	20.3	39.1	47.0
	Iterative RPN+		38.6	58.8	42.2	21.1	41.5	50.0
	GA-RPN		39.5	59.3	43.2	21.8	42.0	50.7
	Cascade RPN		**40.1**	**59.4**	**43.8**	**22.1**	**42.4**	**51.6**
Faster R-CNN	RPN	1000	37.1	**59.3**	40.1	21.4	39.8	46.5
	Cascade RPN		**40.5**	**59.3**	**44.2**	**22.6**	**42.9**	**51.5**
	RPN	300	36.9	58.9	39.9	21.1	39.6	46.5
	Iterative RPN+		39.2	58.2	43.0	21.5	42.0	50.4
	GA-RPN		39.9	**59.4**	43.6	**22.0**	42.6	50.9
	Cascade RPN		**40.6**	58.9	**44.5**	**22.0**	**42.8**	**52.6**