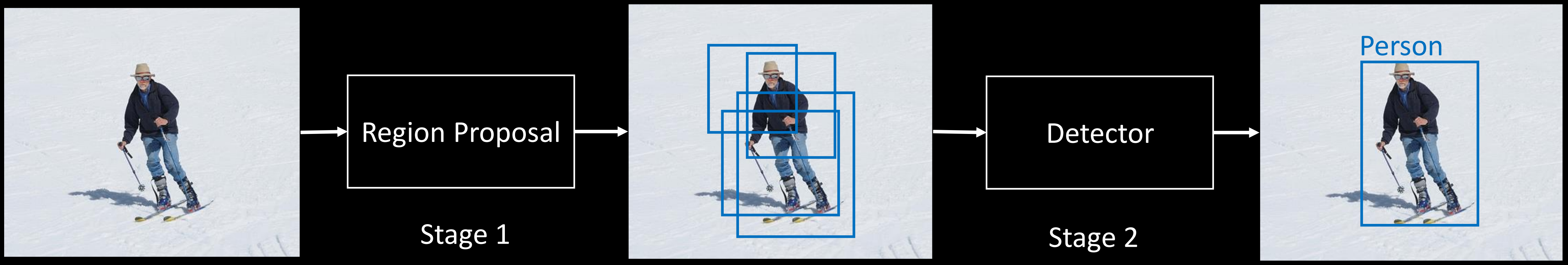
Cascade RPN: Delving into High-Quality Region Proposal Network with Adaptive Convolution

Thang Vu Hyunjun Jang Pham X. Trung Chang D. Yoo

Korea Advanced Institute of Science and Technology

Background


Region Proposal

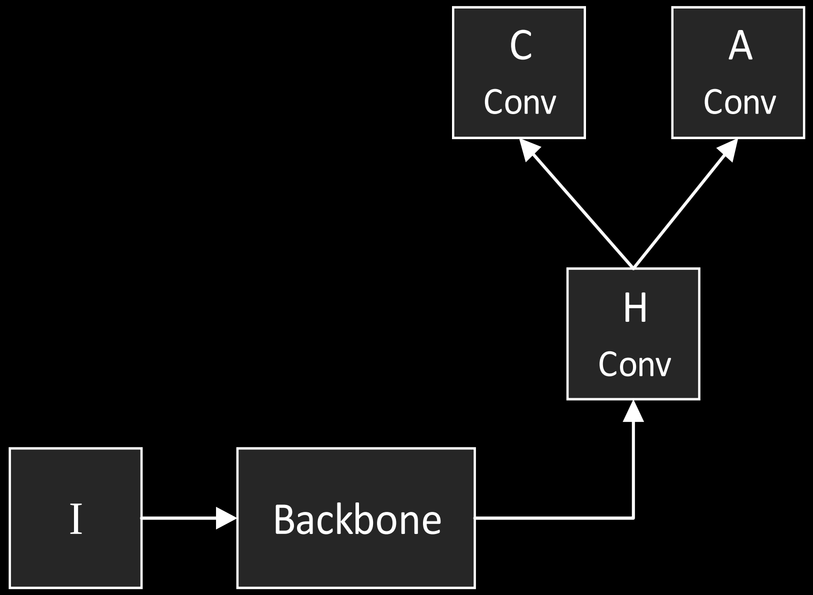
Stage 1

Detector

Stage 2

The proposed method aims to improve the RPN in stage 1

Region proposal network

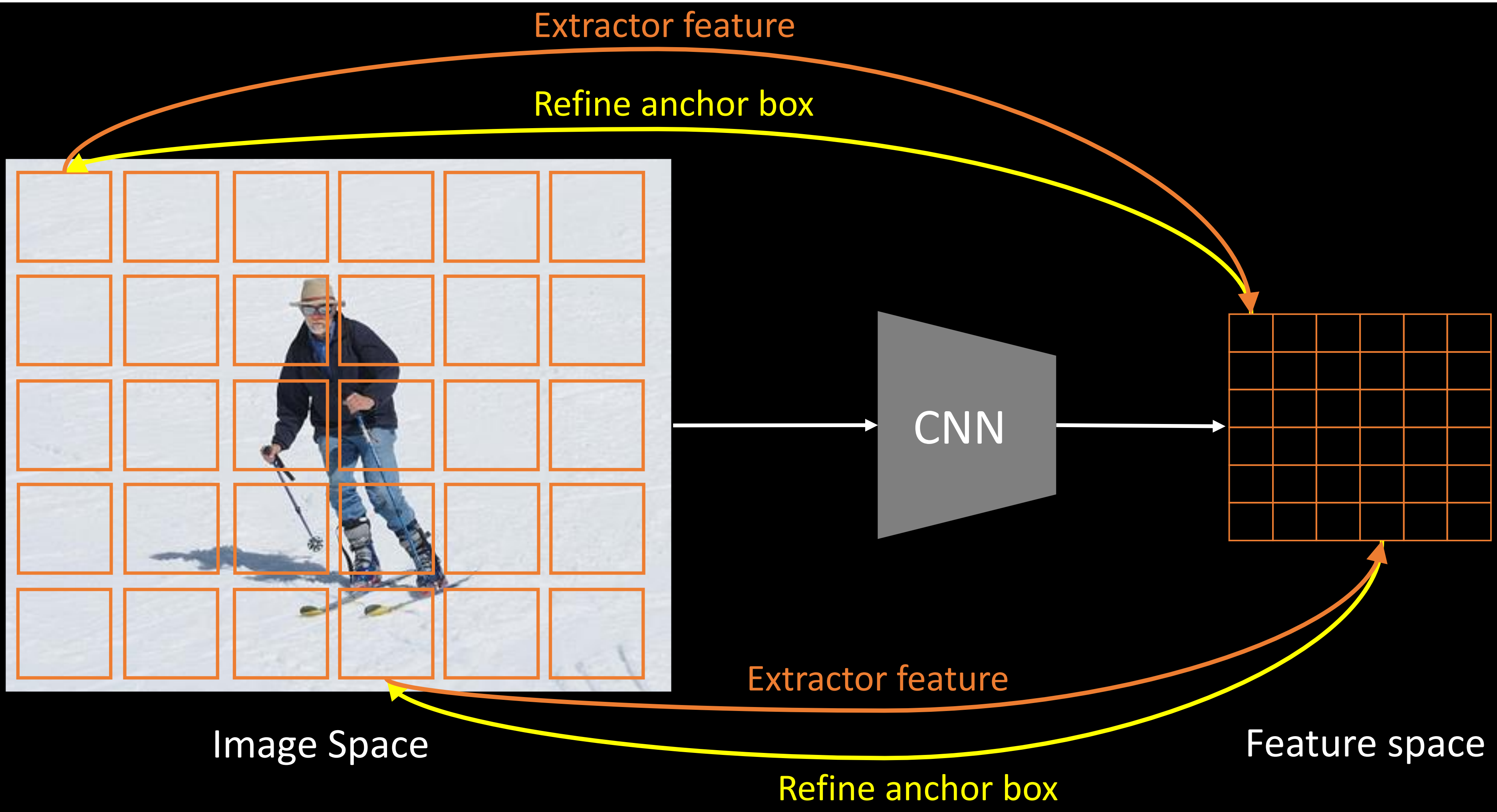


- I: Input image
- Backbone: Feature extractor
- H: Head (shared)
- C: Classifier
- A: Anchor regressor

Region proposal network [1]

[1] Ren et al., Toward real-time object detection with RPN, NeurIPS 2015.

Alignment in RPN



Extractor feature

Refine anchor box

CNN

Extractor feature

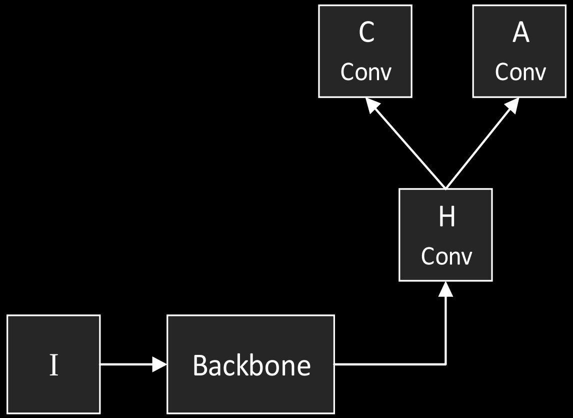
Image Space

Feature space

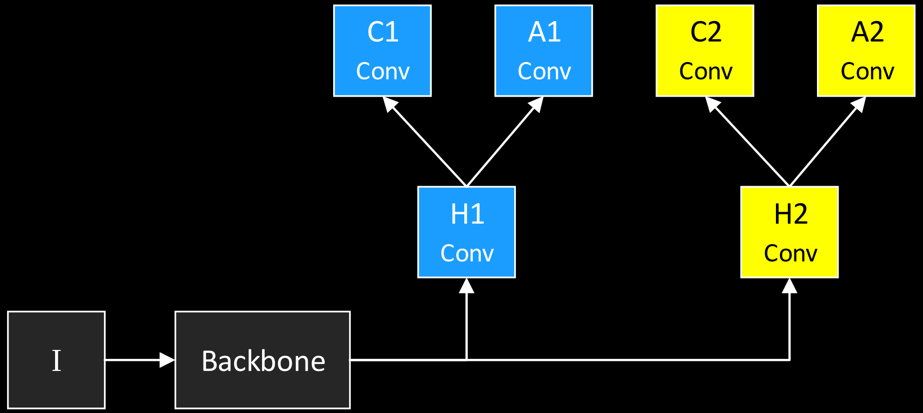
Refine anchor box

Correspondence = Alignment

Iterative RPN



RPN [1]



Iterative RPN [2]

Misalignment

Anchor shape and position change after being refined

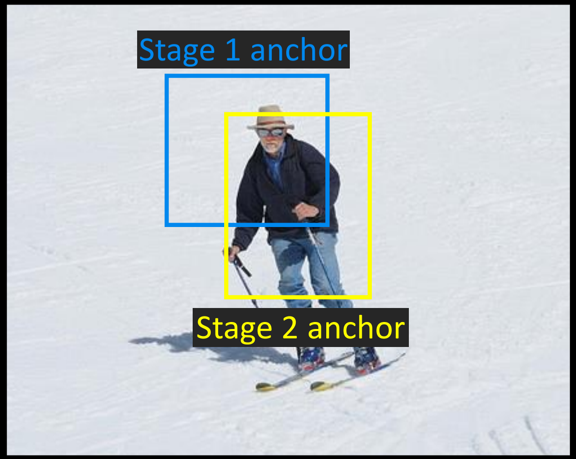
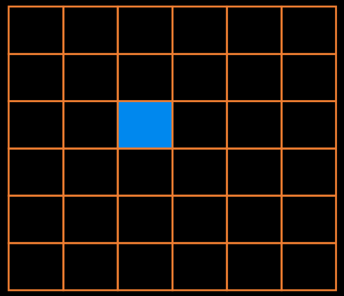


Image space

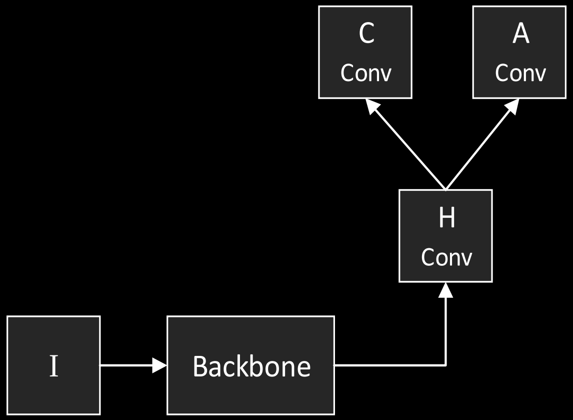


Feature space

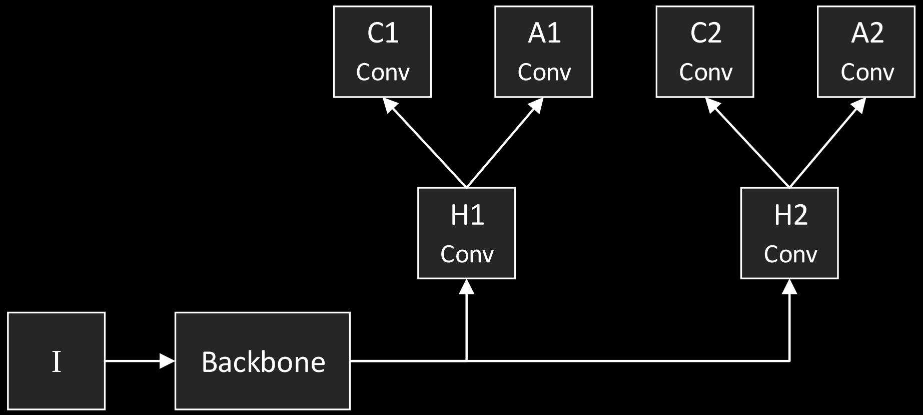
[1] Ren et al., Toward real-time object detection with RPN, NeurIPS 2015.

[2] Zhong et al., Cascade region proposal and global context for deep object detection, arXiv 2018.

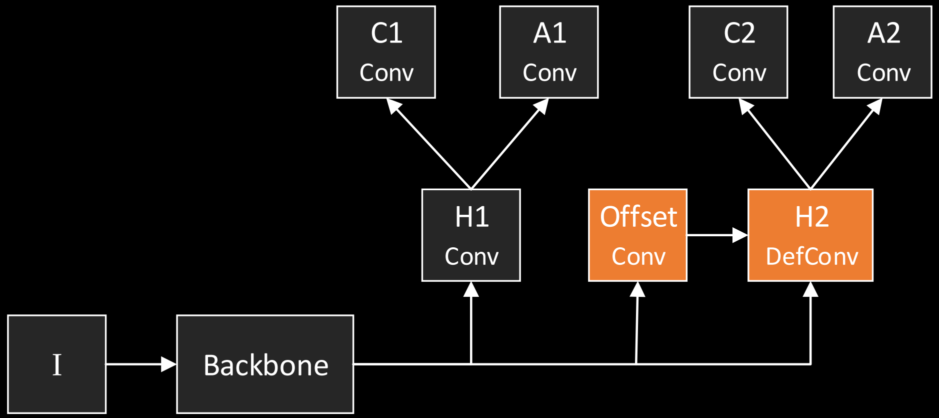
Iterative RPN+ and GA-RPN


C Conv
A Conv
H Conv
I
Backbone

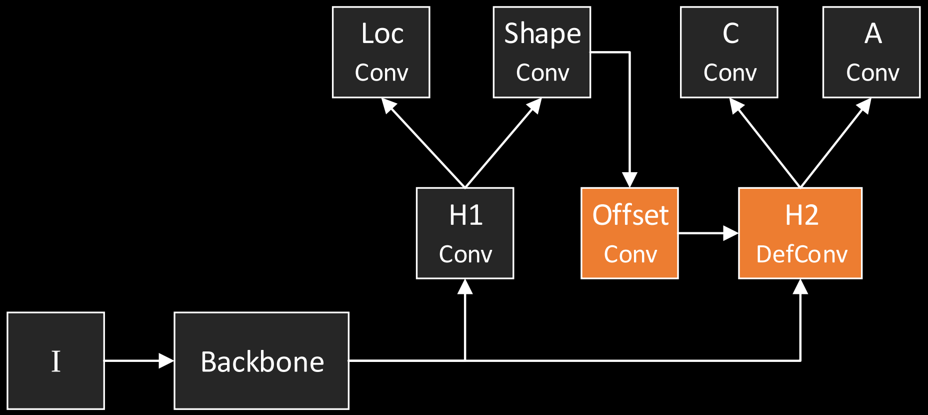
RPN [1]


C1 Conv
A1 Conv
C2 Conv
A2 Conv
H1 Conv
H2 Conv
I
Backbone

Iterative RPN [2]


C1 Conv
A1 Conv
C2 Conv
A2 Conv
H1 Conv
Offset Conv
H2 DefConv
I
Backbone

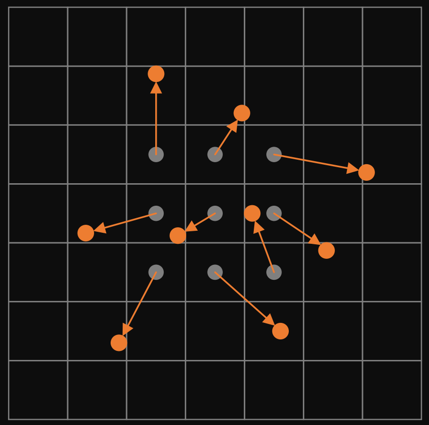
Iterative RPN+ [3]


Loc Conv
Shape Conv
C Conv
A Conv
H1 Conv
Offset Conv
H2 DefConv
I
Backbone

GA-RPN [4]

Misalignment

- Arbitrary feature transform
- No constrains for alignment



Deformable convolution

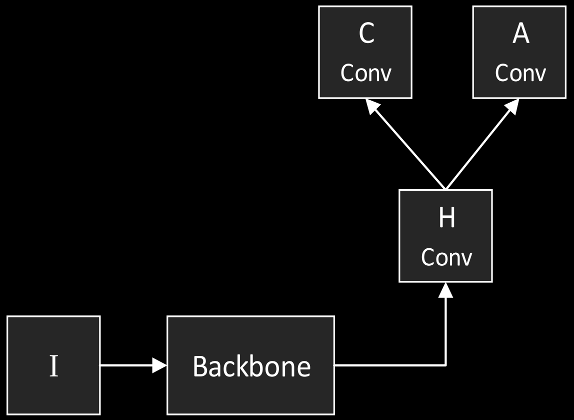
[1] Ren et al., Toward real-time object detection with RPN, NeurIPS 2015.

[2] Zhong et al., Cascade region proposal and global context for deep object detection, arXiv 2018.

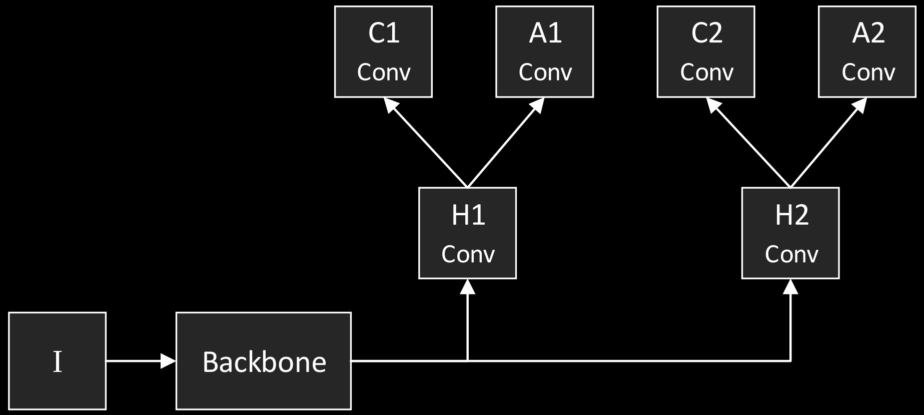
[3] Fan et al., Siamese cascaded region proposal networks for real-time visual tracking. CVPR 2019

[4] Wang et al., Region proposal by guided anchoring, CVPR 2019.

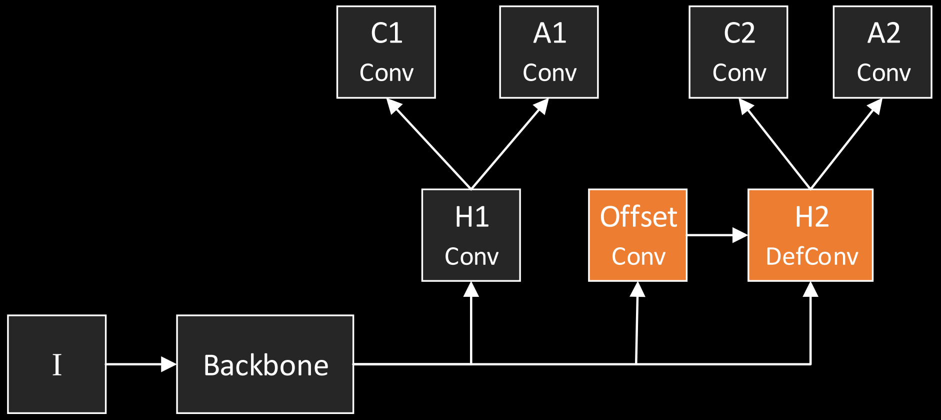
Proposed Cascade RPN



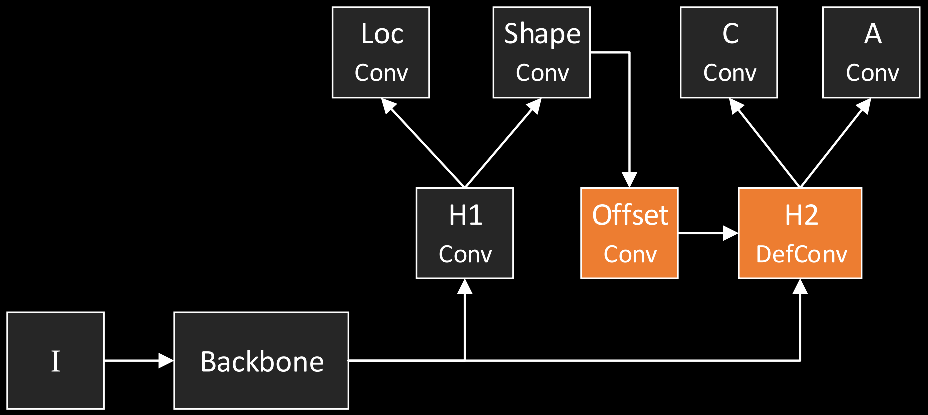
RPN [1]



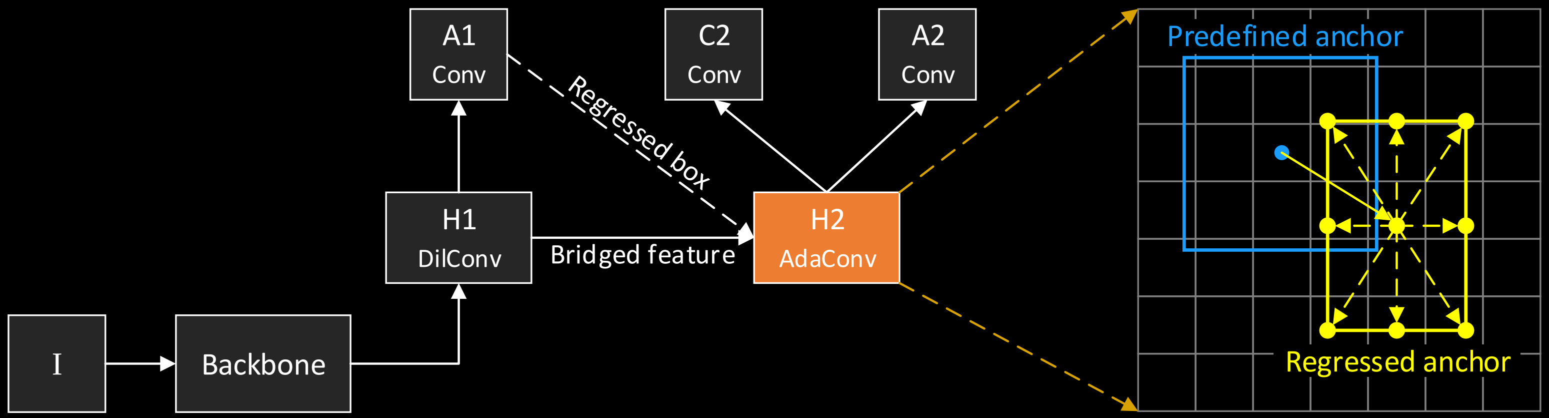
Iterative RPN [2]



Iterative RPN+ [3]



GA-RPN [4]



Cascade RPN (ours)

[1] Ren et al., Toward real-time object detection with RPN, NeurIPS 2015.

[2] Zhong et al., Cascade region proposal and global context for deep object detection, arXiv 2018.

[3] Fan et al., Siamese cascaded region proposal networks for real-time visual tracking. CVPR 2019

[4] Wang et al., Region proposal by guided anchoring, CVPR 2019.

Adaptive Convolution

- Standard Convolution
  - Sample at regular grid $\mathbb{R}$

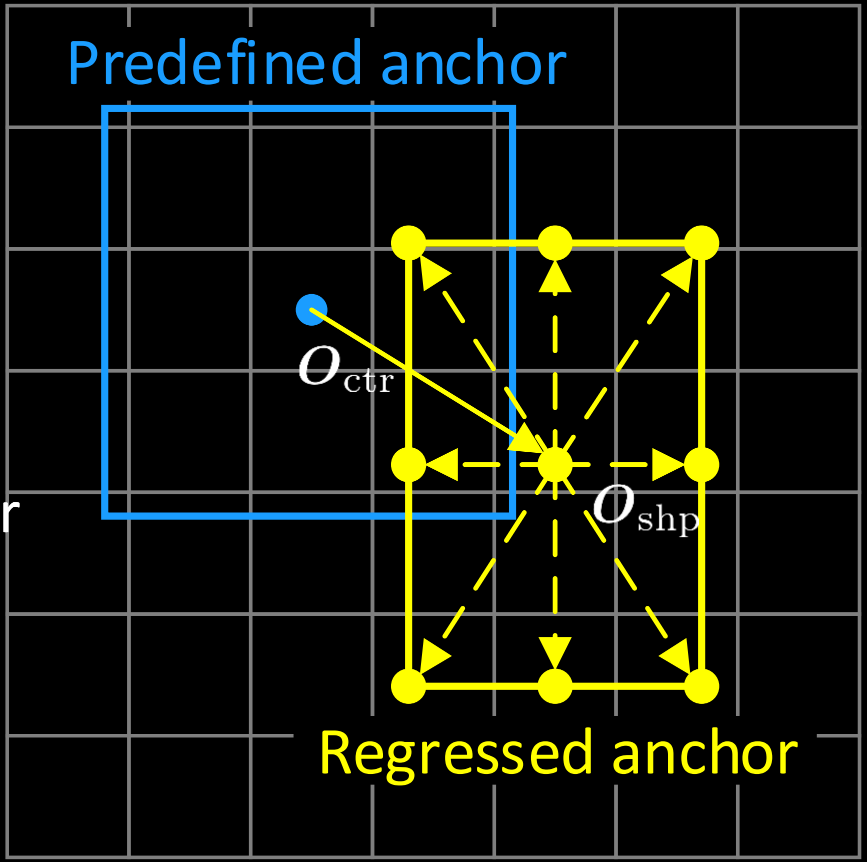
$$\boldsymbol{y}[\boldsymbol{p}] = \boxed{\sum_{\boldsymbol{r}\in\mathbb{R}}} \boldsymbol{w}[\boldsymbol{r}] \cdot \boldsymbol{x}[\boldsymbol{p}+\boldsymbol{r}]$$

$$\mathbb{R} = \{(-1,-1),(-1,0),\ldots,(0,1),(1,1)\}$$

- Adaptive Convolution
  - Sample at offset grid $\mathbb{O}$, guided by anchor

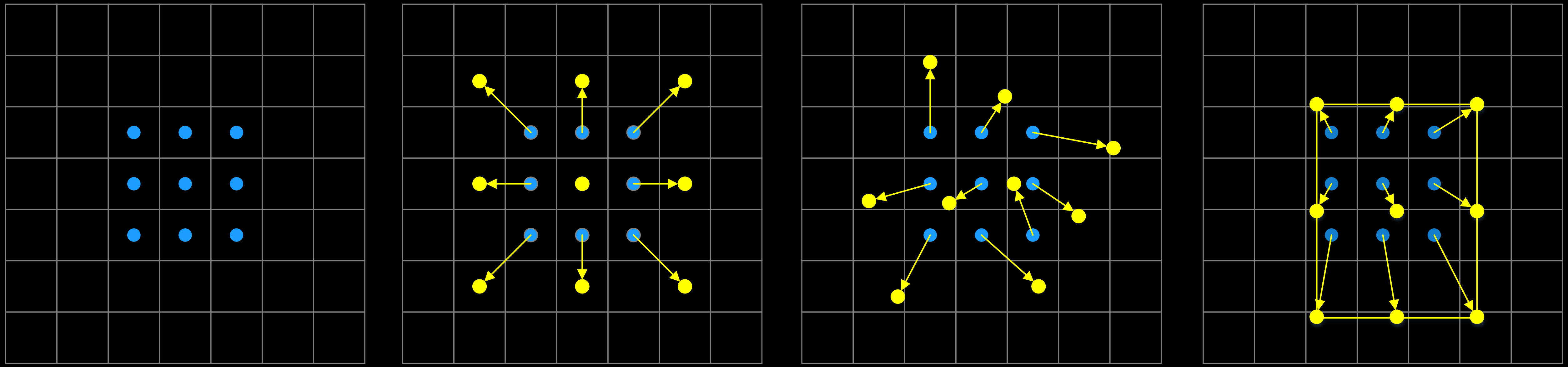
$$\boldsymbol{y}[\boldsymbol{p}] = \boxed{\sum_{\boldsymbol{o}\in\mathbb{O}}} \boldsymbol{w}[\boldsymbol{o}] \cdot \boldsymbol{x}[\boldsymbol{p}+\boldsymbol{o}]$$

$$\boldsymbol{o} = \underset{\substack{\uparrow \\ \text{Position}}}{\boldsymbol{o}_{\mathrm{ctr}}} + \underset{\substack{\uparrow \\ \text{Semantic scope}}}{\boldsymbol{o}_{\mathrm{shp}}}$$



Adaptive conv systematically maintain alignment between features and anchors!

Sampling location



Standard Conv

Dilated Conv [1]

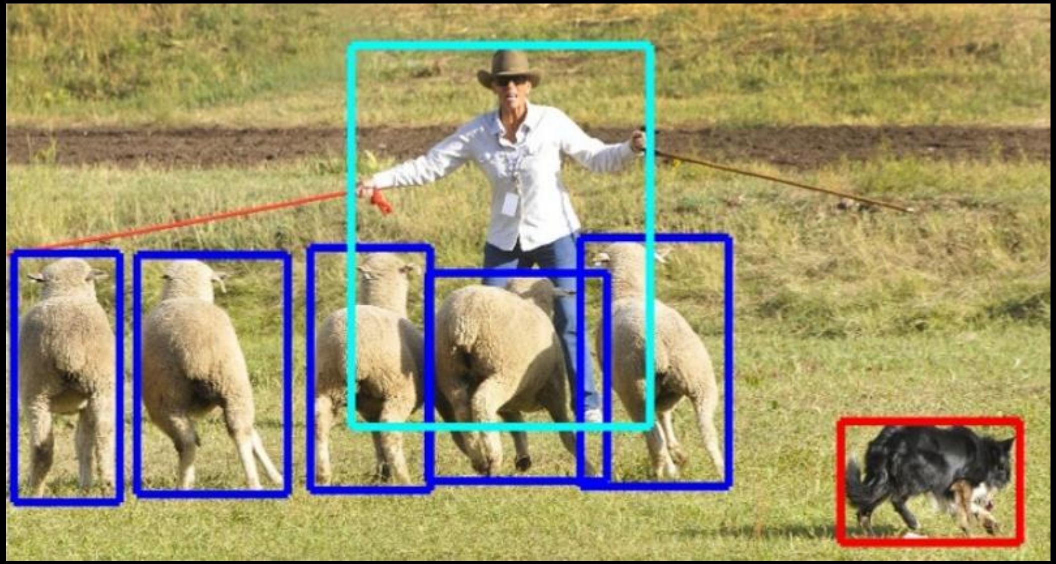
Deformable Conv [2]

Adaptive Conv (ours)

[1] Yu et al. Multi-Scale Context Aggregation by Dilated Convolutions. arXiv 2015.

[2] Dai et al. Deformable Convolutional Networks. ICCV 2017.

Experiments



- Dataset: COCO2017 [1]
  - Train: 115k images
  - Val: 5k images
  - Test-dev: 20k images
- Evaluation metric:
  - Average Recall (AR) for Region Proposal performance
  - Average Precision (AP) for Detection performance
  - Runtime is measured on a single V100

[1] Lin et al. Microsoft COCO: Common Objects in Context, ECCV 2014.

Region Proposal Results

Method	Backbone	AR$_{100}$	AR$_{300}$	AR$_{1000}$	AR$_{S}$	AR$_{M}$	AR$_{L}$	Time (s)
SharpMask [1]	ResNet-50	36.4	-	48.2	-	-	-	0.76
GCN-NS [2]	VGG-16	31.6	-	60.7	-	-	-	0.10
AttractioNet [3]	VGG-16	53.3	-	66.2	31.5	62.2	77.7	4.00
ZIP [4]	BN-inception	53.9	-	76.0	31.9	63.0	78.5	1.13
RPN [5]	ResNet-50	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
Iterativve RPN+		54.0	60.4	63.0	35.6	62.7	73.9	0.06
GA-RPN [6]		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

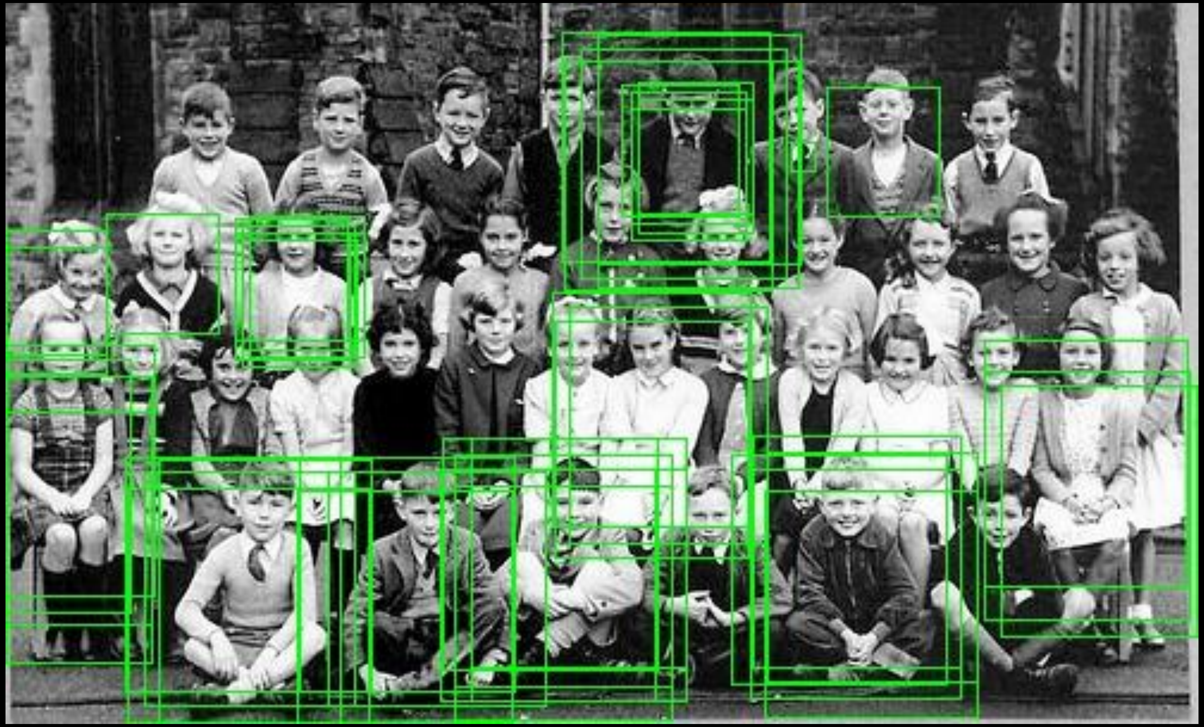
[1] Pinhero et al. Learning to refine object segments. ECCV 2016.
[2] Lu et al. Toward scale-invariance and position-sensitive region proposal networks.. ECCV 2018.
[3] Gidaris et al. Attend refine repeat: Active box proposal generation via in-out localization. arXiv 2016.
[4] Li et al. Zoom out-and-in network with map attention decision for region proposal and object detection. IJCV 2019.
[5] Ren et al. Faster r-cnn: Towards real-time object detection with region proposal networks. NeuIPS 2015.
[6] Wang et al. Region proposal by guided anchoring. CVPR 2019.

Region Proposal Results

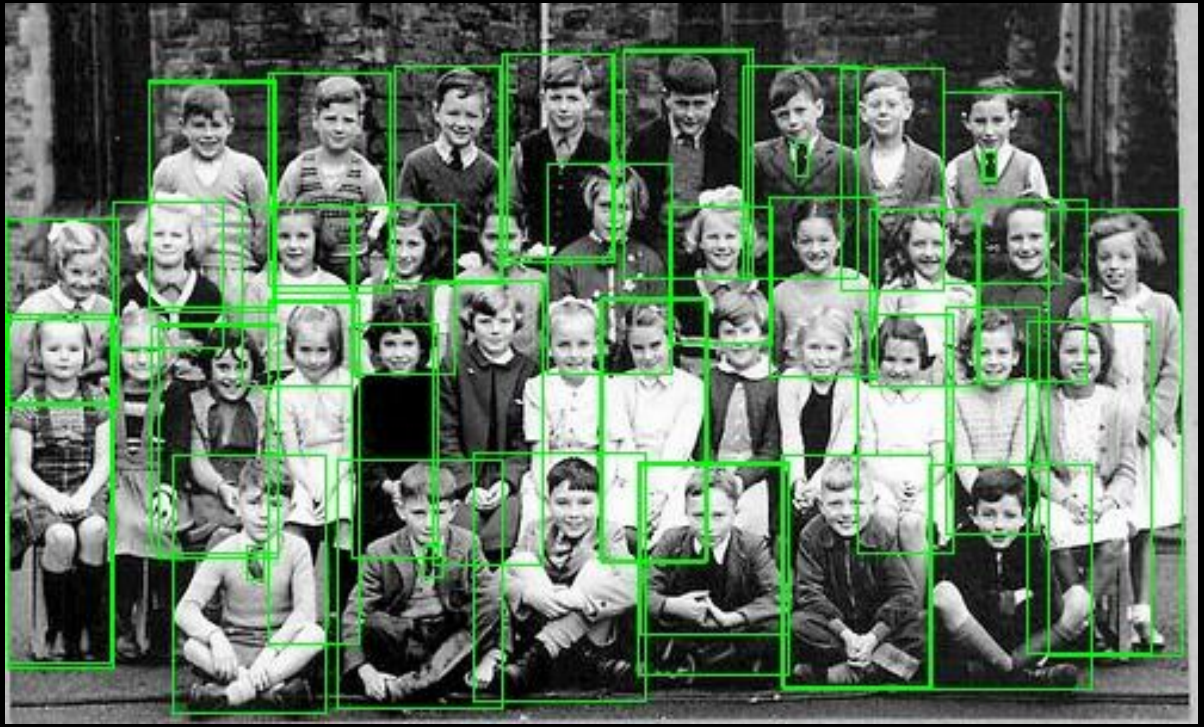
Method	Backbone	AR_{100}	AR_{300}	AR_{1000}	AR_S	AR_M	AR_L	Time (s)
SharpMask [1]	ResNet-50	36.4	-	48.2	-	-	-	0.76
GCN-NS [2]	VGG-16	31.6	-	60.7	-	-	-	0.10
AttractioNet [3]	VGG-16	53.3	-	66.2	31.5	62.2	77.7	4.00
ZIP [4]	BN-inception	53.9	-	76.0	31.9	63.0	78.5	1.13
RPN [5]	ResNet-50	44.6	52.9	58.3	29.5	51.7	61.4	**0.04**
Iterative RPN		48.5	55.4	58.8	32.1	56.9	65.4	0.05
Iterativve RPN+		54.0	60.4	63.0	35.6	62.7	73.9	0.06
GA-RPN [6]		59.1	65.1	68.5	40.7	68.2	78.4	0.06
Cascade RPN		**61.1 (+2.0)**	**67.6 (+2.5)**	**71.7 (+3.2)**	**42.1 (+1.4)**	**69.3 (+1.1)**	**82.8 (+4.4)**	0.06 (+0.0)

[1] Pinhero et al. Learning to refine object segments. ECCV 2016.
[2] Lu et al. Toward scale-invariance and position-sensitive region proposal networks.. ECCV 2018.
[3] Gidaris et al. Attend refine repeat: Active box proposal generation via in-out localization. arXiv 2016.
[4] Li et al. Zoom out-and-in network with map attention decision for region proposal and object detection. IJCV 2019.
[5] Ren et al. Faster r-cnn: Towards real-time object detection with region proposal networks. NeuIPS 2015.
[6] Wang et al. Region proposal by guided anchoring. CVPR 2019.

Qualitative Results

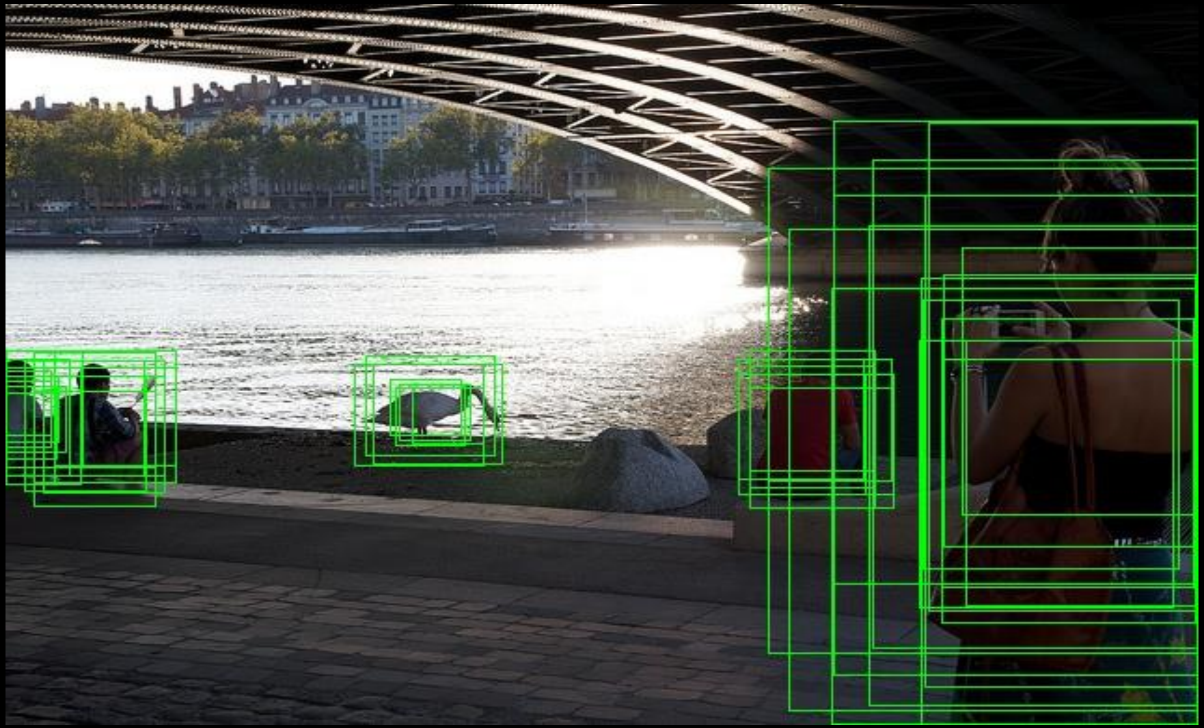


Stage 1

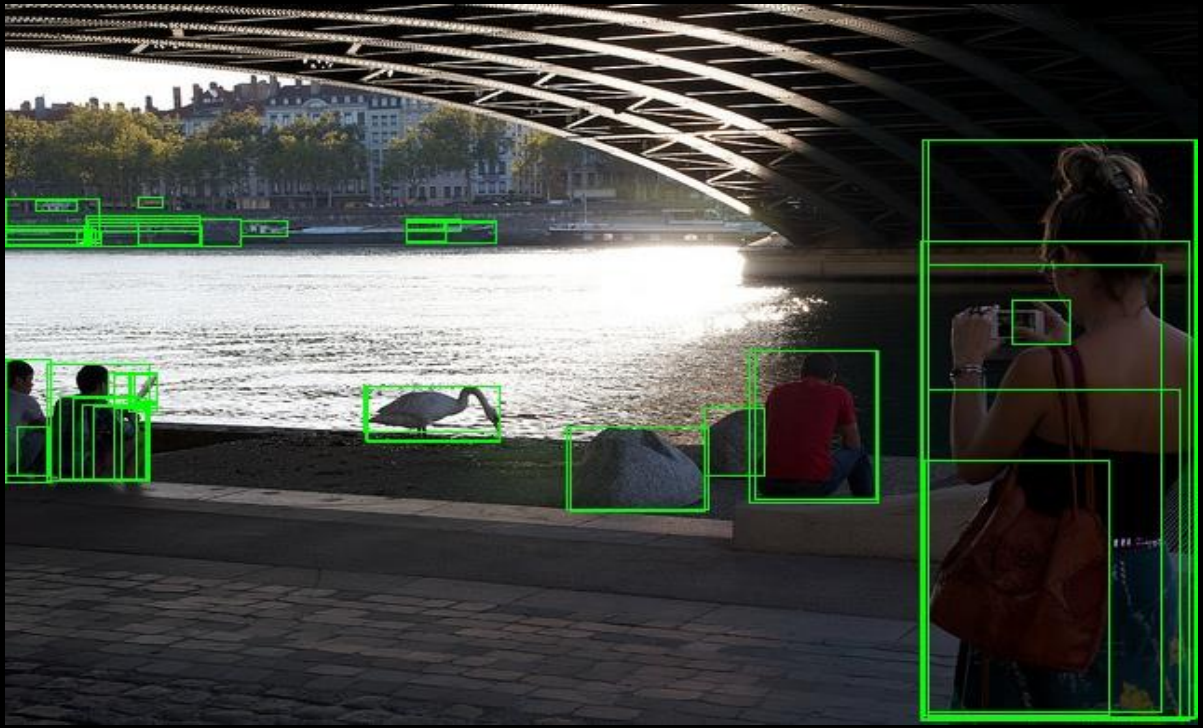


Stage 2

Qualitative Results



Stage 1



Stage 2

Detection Results

Detector	Proposal method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Fast R-CNN [1]	RPN [2]	36.6	58.6	39.5	20.3	39.1	47.0
	Iterative RPN+	38.8	58.8	42.2	21.1	41.5	50.0
	GA-RPN [3]	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 [2]	RPN [2]	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 [3]	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**

[1] Ross B. Girshick. Fast R-CNN. ICCV 2015.

[2] Ren et al. Faster r-cnn: Towards real-time object detection with region proposal networks. NeuIPS 2015.

[3] Wang et al. Region proposal by guided anchoring. CVPR 2019.

Summary

- Alignment is not well persevered in existing multi-stage RPN.
- Cascade RPN systematically ensures alignment by Adaptive Convolution.
- Cascade RPN achieves state-of the-art proposal performance on COCO dataset.

Code is available at:
https://github.com/thangvubk/Cascade-RPN

Poster #86 at East Exhibition Hall B + C

Thank you!