Improving Proposal-Based Object Detection Using Convolutional Context Features

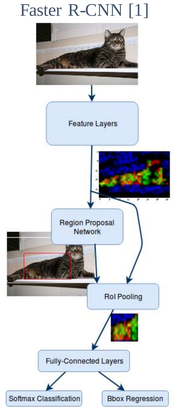
Emre Can Kaya
e174127@metu.edu.tr

A. Aydın Alatan
alatan@metu.edu.tr

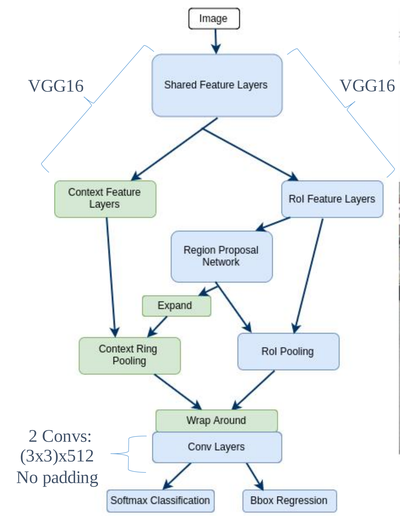
Motivation

- In proposal-based detection, a common practice is that once region proposals are determined, information coming from features surrounding the proposals is left out when inferring final detections.
- In this study, it is argued that local context of a region of interest can be further exploited by learning separate context features.

Faster R-CNN [1]

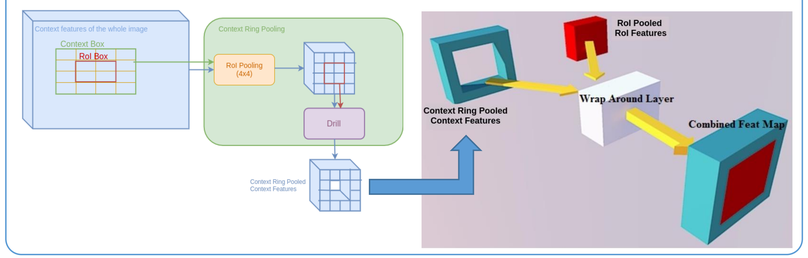


Proposed Method

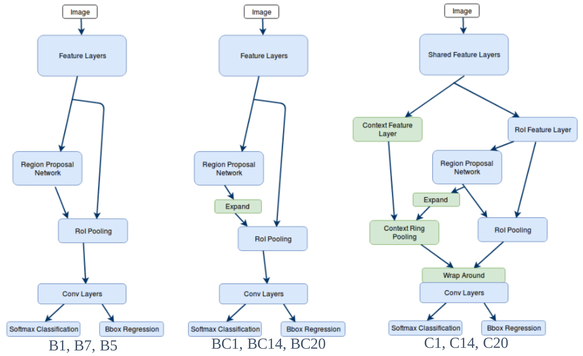


Offset Ratio = 

$$Offset\ Ratio = \frac{(contextpool\ size) - (roipool\ size)}{2 * (roipool\ size)}$$



Experimental Work



Results are on PASCAL VOC 2007

The requirement for learning separate features for the context

Model	roipool(/ctxpool)	# of Ctx Layers	Offset Ratio
B_1	6x6	0	-
BC_1	8x8	0	0.17
C_1	6x6/8x8	1	0.17
B_7	7x7	0	-
BC_{14}	9x9	0	0.14
C_{14}	7x7/9x9	1	0.14
B_5	5x5	0	-
BC_{20}	7x7	0	0.20
C_{20}	5x5/7x7	1	0.20

Class	B_1	BC_1	C_1	B_7	BC_{14}	C_{14}	B_5	BC_{20}	C_{20}
a.plane	58.94	61.29	**67.67**	**60.58**	59.84	59.79	58.56	62.29	**63.36**
bicycle	69.78	73.34	**76.42**	71.83	**73.88**	73.69	72.12	72.94	**73.35**
bird	55.60	55.46	**61.13**	56.45	54.88	**59.70**	55.22	54.09	**58.68**
boat	38.54	45.32	**54.85**	39.07	38.87	**41.93**	39.46	40.47	**41.37**
bottle	30.10	33.10	**34.09**	27.03	29.41	**29.56**	28.25	31.23	**32.27**
bus	67.96	69.78	**73.81**	**68.08**	65.73	67.71	64.87	**71.92**	69.68
car	68.28	67.96	**74.68**	67.45	66.95	**70.93**	67.22	67.76	**70.80**
cat	74.46	75.06	**80.32**	73.28	73.31	**77.85**	73.58	75.07	**76.06**
chair	38.15	37.44	**40.30**	36.79	**37.26**	37.04	35.38	35.47	**38.36**
cow	61.37	62.26	**71.13**	62.04	63.95	**64.46**	58.00	64.10	**64.45**
d.table	55.46	58.01	**59.75**	56.07	**58.61**	56.75	52.49	**61.46**	58.22
dog	72.91	74.08	**76.35**	71.24	71.17	**74.11**	72.90	72.09	**75.04**
horse	75.23	74.91	**79.36**	76.32	75.71	**76.68**	74.66	76.63	**77.27**
m.bike	70.00	71.48	**75.84**	71.93	69.79	**73.11**	66.26	68.68	**71.67**
person	64.93	64.33	**68.11**	64.38	63.70	**66.26**	63.57	64.36	**66.31**
p.plant	27.59	27.15	**31.13**	**31.00**	29.50	29.88	26.03	26.29	**28.81**
sheep	54.24	53.37	**63.26**	52.89	**54.52**	51.84	48.93	**55.58**	54.61
sofa	60.08	57.88	**62.65**	59.64	56.41	**60.95**	56.31	58.37	**63.71**
train	72.35	70.47	**73.63**	70.86	68.60	**71.61**	70.97	71.05	**73.51**
tv	**62.80**	60.88	62.48	59.38	60.67	**61.75**	61.28	60.63	**62.24**
mean	58.94	59.68	**64.35**	58.82	58.64	**60.28**	57.30	59.52	**60.99**

Performance with different number of Ctx/RoI layers

Model	roipool(/ctxpool)	# of Ctx Layers
B_1	6x6	0
C_1	6x6/8x8	1
C_{f2}	6x6/8x8	2
C_{f3}	6x6/8x8	3

Class	B_1	C_1	C_{f2}	C_{f3}
aeroplane	58.94	**67.67**	60.58	61.20
bicycle	69.78	**76.42**	70.28	69.61
bird	55.60	**61.13**	56.18	57.00
boat	38.54	**54.85**	40.25	42.44
bottle	30.10	34.09	**35.36**	30.62
bus	67.96	**73.81**	67.45	70.36
car	68.28	**74.68**	69.19	70.77
cat	74.46	**80.32**	74.41	77.47
chair	38.15	**40.30**	34.68	36.03
cow	61.37	**71.13**	62.07	64.42
diningtable	55.46	58.01	**59.75**	56.73
dog	72.91	**76.35**	71.33	72.23
horse	75.23	**79.36**	76.65	75.89
motorbike	70.00	**75.84**	68.87	70.83
person	64.93	**68.11**	64.88	65.52
pottedplant	27.59	**31.13**	26.41	29.38
sheep	54.24	**63.26**	51.91	53.40
sofa	60.08	62.65	59.72	**63.33**
train	72.35	**73.63**	73.61	73.44
tv/monitor	62.80	62.48	**64.15**	62.42
mean	58.94	**64.35**	59.24	60.18

Effect of Offset Ratio

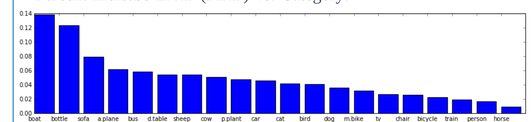
Model	roipool(/ctxpool)	# of Ctx Layers	Offset Ratio
B_1	6x6	0	-
C_1	6x6/8x8	1	0.17
C_{33}	6x6/10x10	1	0.33
B_7	7x7	0	-
C_{14}	7x7/9x9	1	0.14
C_{29}	7x7/11x11	1	0.29
B_5	5x5	0	-
C_{20}	5x5/7x7	1	0.20
B_8	8x8	0	-
C_{12}	8x8/10x10	1	0.125
C_{25}	8x8/12x12	1	0.250
B_9	9x9	0	-
C_{22}	9x9/13x13	1	0.22

Class	B_1	C_1	C_{33}	B_7	C_{14}	C_{29}	B_5	C_{20}	B_8	C_{12}	C_{25}	B_9	C_{22}
aeroplane	58.94	**67.67**	57.95	60.58	59.79	**66.03**	58.56	**63.36**	58.47	60.69	**65.12**	59.61	**62.76**
bicycle	69.78	**76.42**	69.59	71.83	73.69	**73.84**	72.12	**73.35**	72.36	71.43	**72.55**	71.22	**72.90**
bird	55.60	**61.13**	53.71	56.45	59.70	**61.16**	55.22	**58.68**	56.23	**58.36**	55.80	52.75	**54.17**
boat	38.54	**54.85**	39.73	39.07	41.93	**48.34**	39.46	**41.37**	41.20	41.32	**42.06**	35.21	**44.86**
bottle	30.10	**34.09**	29.19	27.03	29.56	**35.11**	28.25	**32.27**	29.05	32.90	**33.13**	31.85	**34.34**
bus	67.96	**73.81**	69.97	68.08	67.71	**74.03**	64.87	**69.68**	64.29	66.69	**69.52**	66.34	**71.58**
car	68.28	**74.68**	68.07	67.45	70.93	**74.55**	67.22	**70.80**	68.80	69.92	**71.04**	68.80	**70.12**
cat	74.46	**80.32**	73.50	73.28	77.85	**80.62**	73.58	**76.06**	74.83	75.72	**75.94**	74.07	**77.28**
chair	38.15	**40.30**	34.02	36.79	37.04	**39.79**	35.38	**38.36**	36.97	36.70	**36.99**	34.18	**37.25**
cow	61.37	**71.13**	60.18	62.04	64.46	**67.45**	58.00	**64.45**	62.46	61.73	60.85	59.02	**63.12**
diningtable	55.46	**59.75**	50.29	56.07	56.75	**63.40**	52.49	**58.22**	55.34	57.23	**58.36**	53.44	**59.42**
dog	72.91	**76.35**	70.40	71.24	74.11	**76.40**	72.90	**75.04**	69.71	72.82	**73.28**	71.04	**73.43**
horse	75.23	**79.36**	72.70	76.32	76.68	**78.16**	74.66	**77.27**	**76.04**	75.14	75.23	76.16	**77.01**
motorbike	70.00	**75.84**	68.89	71.93	73.11	**74.78**	66.26	**71.67**	69.75	**72.77**	71.30	**71.11**	69.79
person	64.93	**68.11**	61.43	64.38	66.26	**67.68**	63.57	**66.31**	65.48	**65.96**	65.89	65.41	**65.55**
pottedplant	27.59	**31.13**	27.89	**31.00**	29.88	29.59	26.03	**28.81**	27.11	29.07	**30.09**	27.35	**28.31**
sheep	54.24	**63.26**	52.55	52.89	51.84	**61.49**	48.93	**54.61**	53.86	**53.93**	52.29	52.12	**55.53**
sofa	60.08	62.65	**64.50**	59.64	60.95	**65.00**	56.31	**63.71**	56.63	63.17	**65.22**	59.83	**60.17**
train	72.35	**73.63**	67.63	70.86	71.61	**75.51**	70.97	**73.51**	70.38	72.63	**73.86**	70.77	**71.26**
tv/monitor	**62.80**	62.48	61.22	59.38	61.75	**64.86**	61.28	**62.24**	60.82	62.79	**63.88**	62.78	**63.42**
mean	58.94	**64.35**	57.67	58.82	60.28	**63.89**	57.30	**60.99**	58.49	60.05	**60.62**	58.15	**60.61**
Offset Ratio	-	0.17	0.33	-	0.14	0.29	-	0.20	-	0.125	0.25	-	0.22

Comparison with Original Faster R-CNN

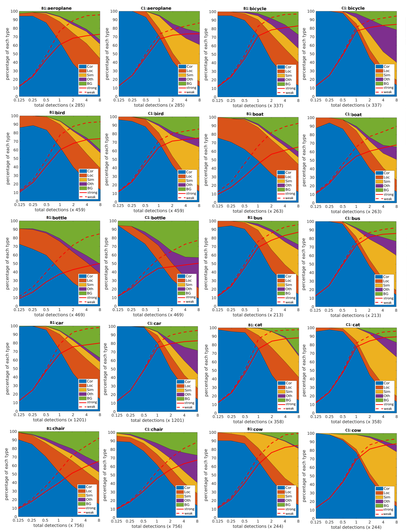
$$PIAP(C) = \frac{AP(C) - BaselineAP(C)}{BaselineAP(C)}$$

Percent Increase in AP (PIAP) vs. Category:

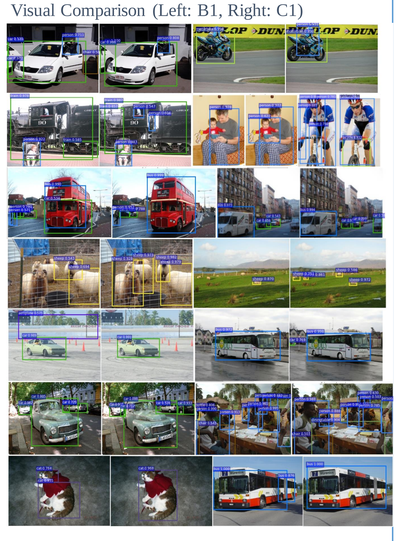


Class	FRCNN [1]	C_1(ours)
aeroplane	**70.0**	67.7
bicycle	**80.6**	76.4
bird	**70.1**	61.1
boat	**57.3**	54.9
bottle	**49.9**	34.1
bus	**78.2**	73.8
car	**80.4**	74.7
cat	**82.0**	80.3
chair	**52.2**	40.3
cow	**75.3**	71.1
diningtable	**67.2**	58.0
dog	**80.3**	76.4
horse	**79.8**	79.4
motorbike	75.0	**75.8**
person	**76.3**	68.1
pottedplant	**39.1**	31.1
sheep	**68.3**	63.3
sofa	**67.3**	62.7
train	**81.1**	73.6
tv/monitor	**67.6**	62.5
mean	**69.9**	64.4
# net params (in Millions)	137.08	25.04

False Positives Comparison [2] (Left: B1, Right: C1)



Visual Comparison (Left: B1, Right: C1)



Conclusion

- Context improves object detection performance
- A separate feature extractor for the context is essential.
- Number of context/RoI feature layers and offset ratio significantly affect the performance
- Higher improvements are observed for categories having distinctive context.
- Usage of context features mainly improves localization. It sometimes acts as a source of confusion between different categories.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91–99.

[2] D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors," in European conference on computer vision. Springer, 2012, pp. 340–353.