Small Object Detection by Generative and Discriminative Learning

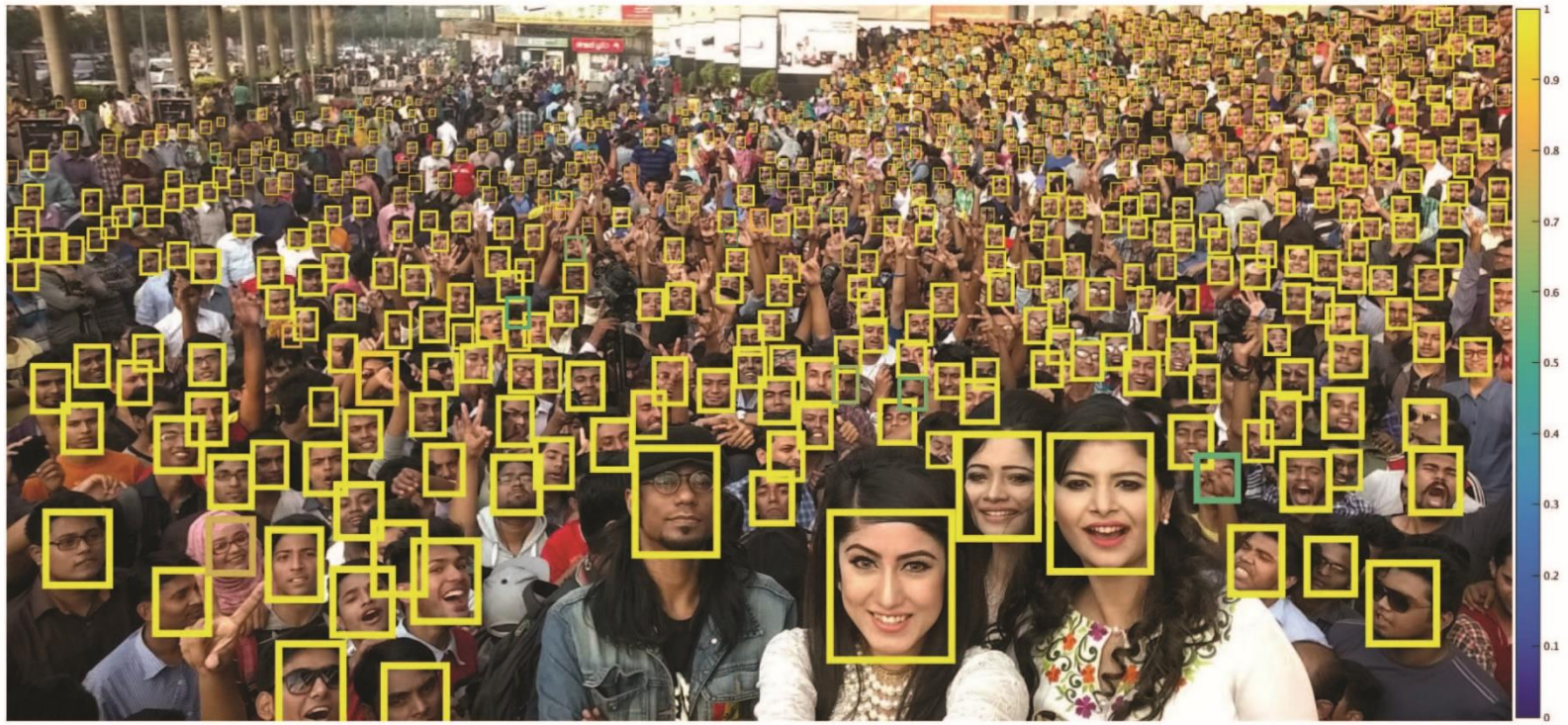
Yi Gu, Jie Li*, Chentao Wu, Weijia Jia and Jianping Chen

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Contents

Small Object Detection

- Convolutional neural network (CNN) has made great advances on object detection
- Model performance on small object is still worse than that of medium and large ones

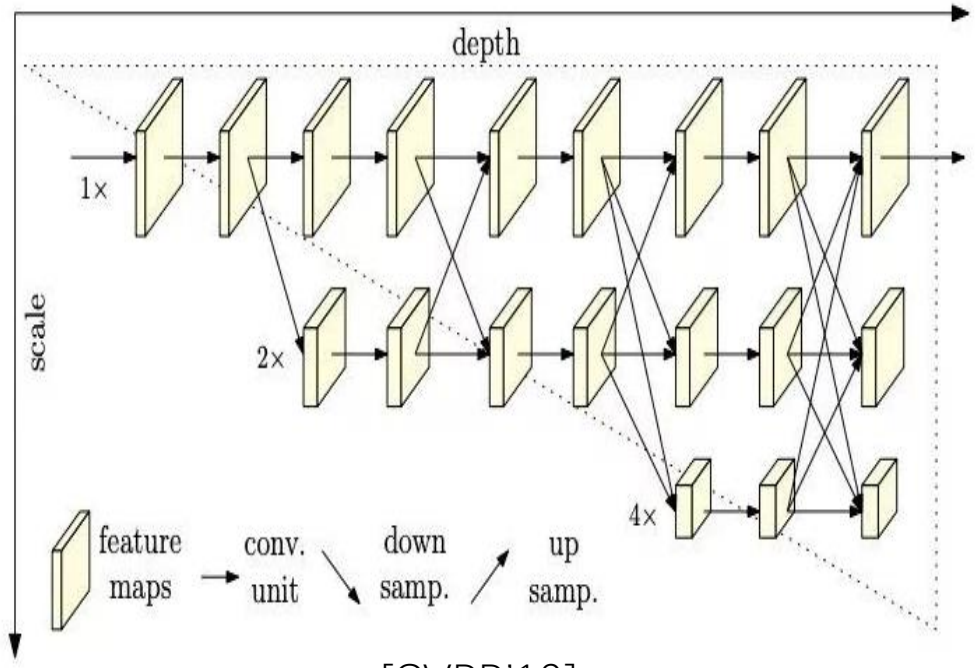


[CVPR'17]

Related Work

Up Scale [NIPS'15, NIPS'16, PAMI'16]

- Reduce input resolution
- Make image more blurred
- Feature information remains unchanged

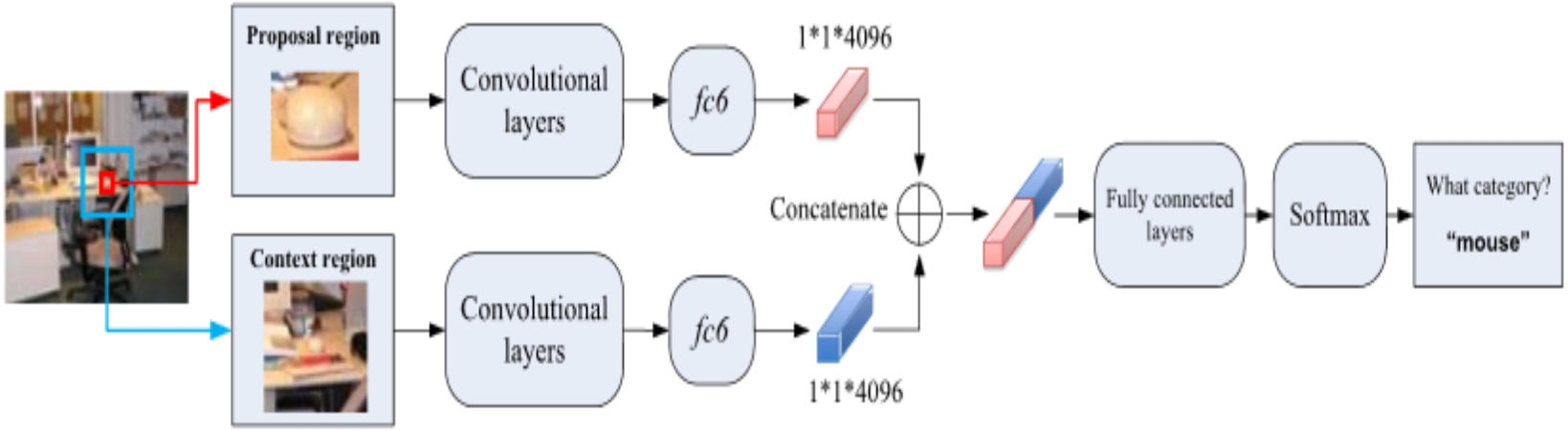


[CVPR'19]

Related Work

Contextual Information [CVPR'15, ACCV'16]

- Utilize context surrounding small objects to fuel information
- Requirement for sufficiently useful surrounding context
- Feature map may contain more interference

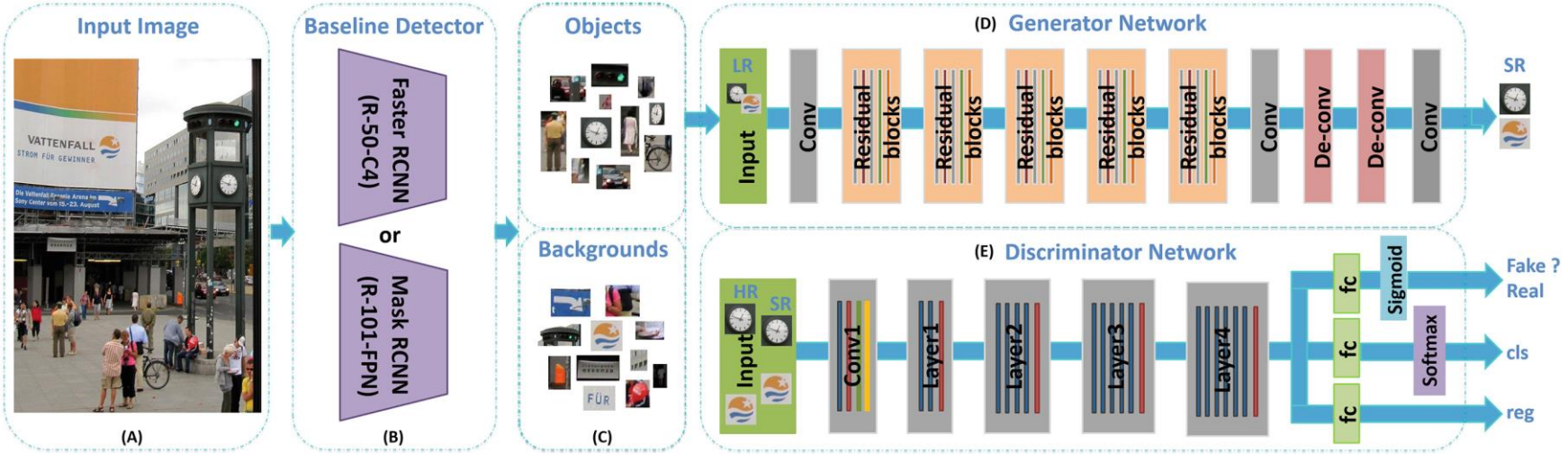


[ACCV'16]

Related Work

Super Resolution [CVPR'17, CVPR'18, ECCV'18, PR'19]

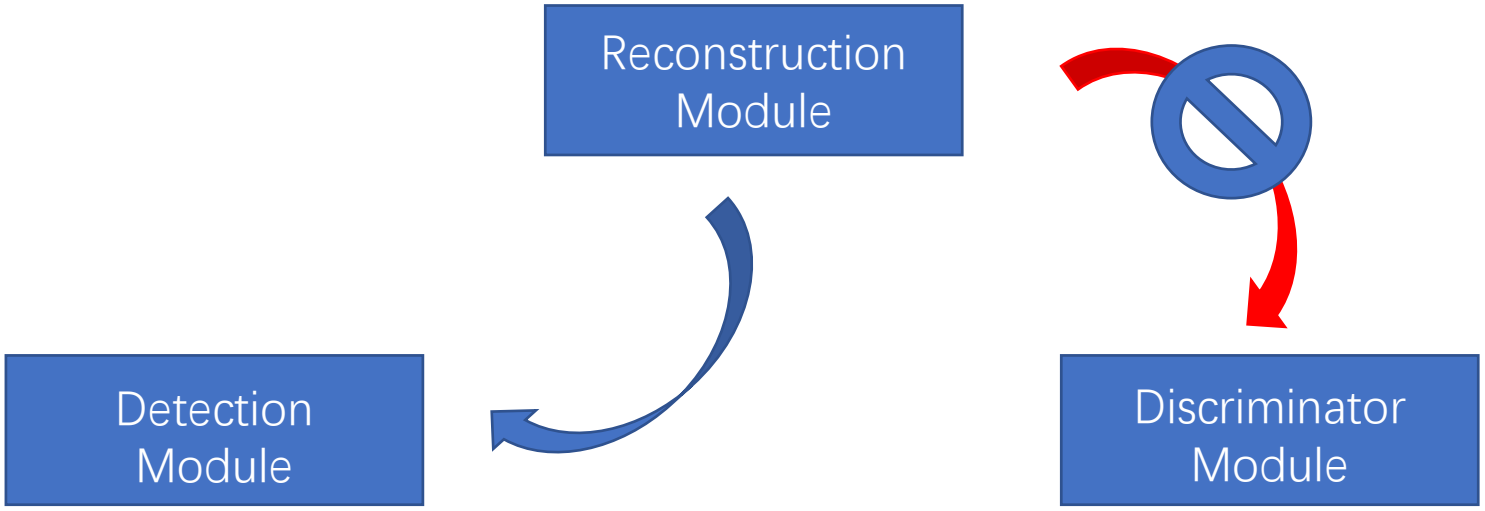
- Implement generative network to increase bounding box resolution
- Singly improve classification accuracy
- Predict bounding box with limited object pixels.



[ECCV'18]

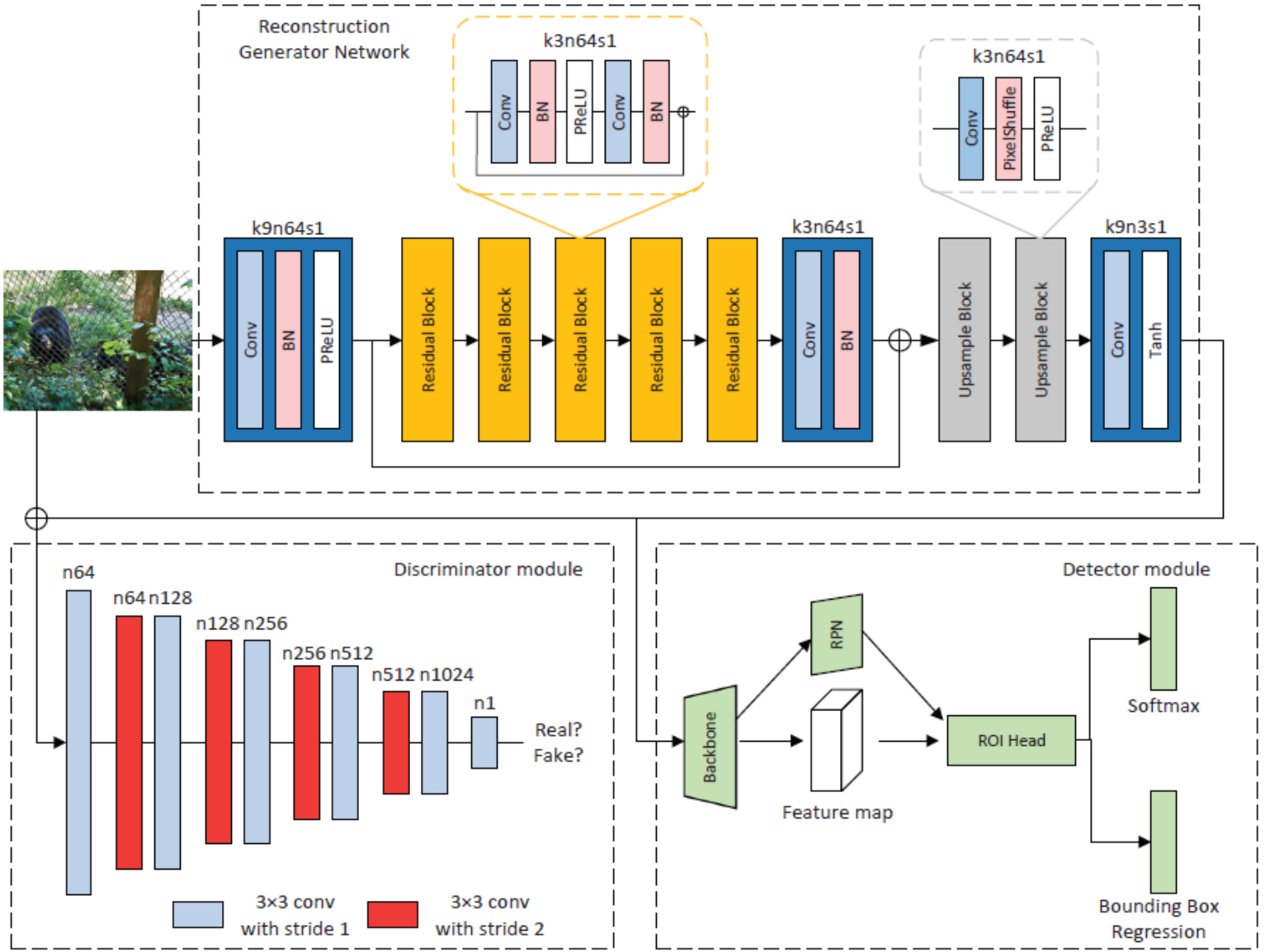
Motivation

- Prior: All three methods apply reconstruction to the discriminator module for small object classification
- Bottleneck: The accuracy of predicting bounding box
- Our motivation: Implement a reconstruction network on the detection module to produce more details for bounding box prediction.



Contents

Our Framework



Loss Function

- $L_{adv} = \frac{1}{m}\sum_{i=1}^{m} \log(1 - D_{\theta_D}(G_{\theta_G}(z_{LR}^{(i)})))$
- $L_{MSE} = \frac{1}{m}\sum_{i=1}^{m} \left\| z_{HR}^{(i)} - G_{\theta_G}(z_{LR}^{(i)}) \right\|_2^2$
- $L_{TV} = \frac{1}{r^2WH}\sum_{i=1}^{rW}\sum_{j=1}^{rH} \left\| \nabla G_{\theta_G}(z_{LR}^{(i,j)}) \right\|$
- $L_{cls} = \frac{1}{m}\sum_{i=1}^{m}[-\left(\log\left(p^{(i)}p^{*(i)} + \left(1 - p^{(i)}\right)\left(1 - p^{*(i)}\right)\right) - \log(D_{cls}\left(G_{\theta_G}\left(z_{LR}^{(i)}\right)\right)\right) + \log\left(D_{cls}\left(z_{HR}^{(i)}\right)\right)]$
- $L_{reg} = \frac{1}{m}\sum_{i=1}^{m}\sum_{j\in(x,y,w,h)}[u_i \geq 1]\left(S_{L_1}\left(t_{SR}^{i,j} - t^{*(i,j)}\right)\right)$
- $\max_{\theta_D}\min_{\theta_G}(\frac{1}{m}\sum_{i=1}^{m}\log D_{\theta_D}(z_{HR}^{(i)})) + \alpha\, L_{adv} + \beta\, L_{cls} + \gamma\, L_{reg} + L_{TV} + L_{MSE}$

Contents

Quantitative Results

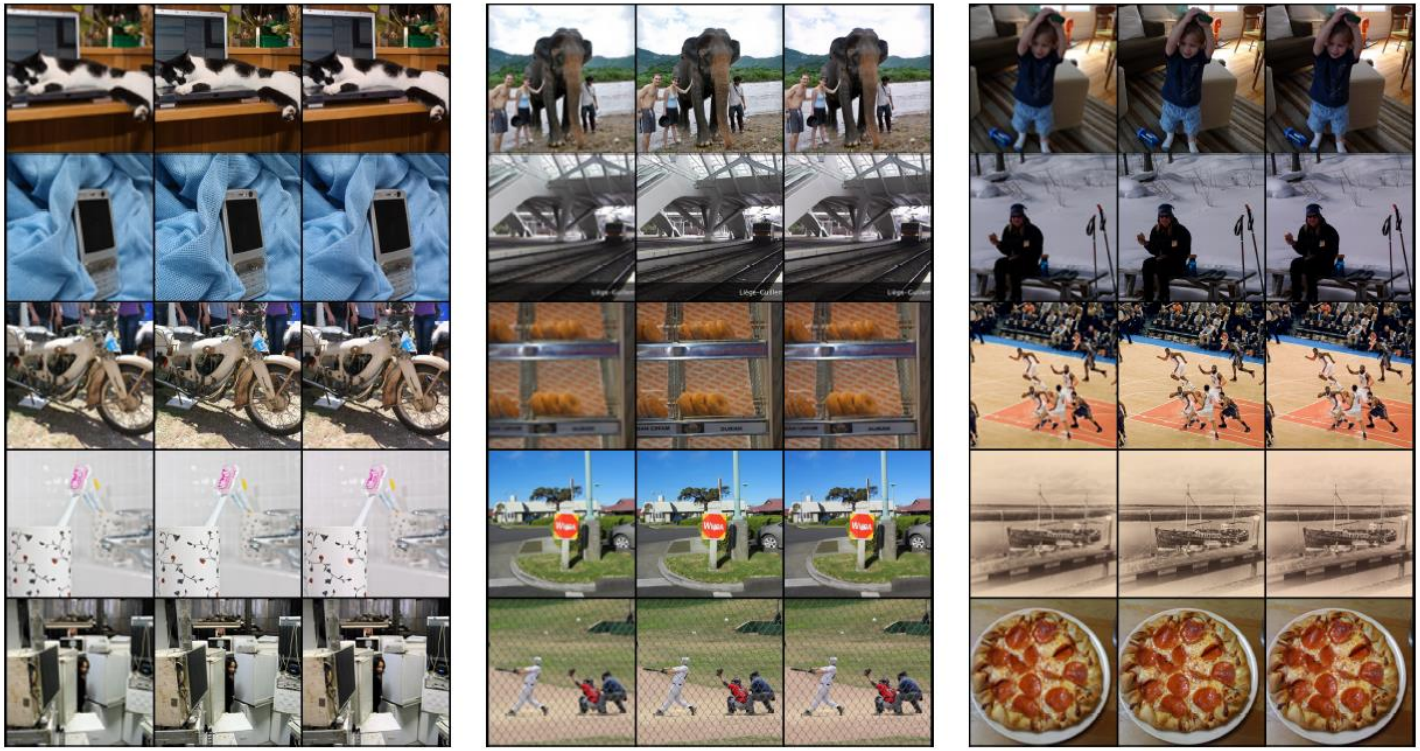
General Object Detection Models

Methods	Backbone	Avg. Precision, IoU:			Avg. Precision, Area:		
		0.5:0.95	0.5	0.75	S	M	L
Faster R-CNN	ResNet-50	31.8	52.1	33.7	15.5	36.9	43.3
Faster R-CNN	ResNet-50-FPN	34.2	55.0	36.4	20.0	37.5	43.9
Faster R-CNN	ResNet-101	35.0	55.1	37.7	17.3	39.8	48.4
Faster R-CNN	ResNet-101-FPN	36.2	57.3	39.2	21.3	39.8	46.6
RetinaNet	ResNet-50-FPN	32.6	50.8	35.0	18.6	35.9	42.4
RetinaNet	ResNet-101-FPN	35.0	53.9	37.7	19.2	38.9	45.7
Mask R-CNN	ResNet-50	32.9	52.7	35.1	17.0	37.5	44.7
Mask R-CNN	ResNet-50-FPN	34.7	55.1	37.7	19.5	37.8	44.9
Mask R-CNN	ResNet-101	36.1	55.5	39.0	18.5	40.9	49.9
Mask R-CNN	ResNet-101-FPN	37.3	58.1	40.7	21.7	40.8	48.1
PanopticFPN	ResNet-50	33.5	54.1	36.2	19.3	36.4	43.0
PanopticFPN	ResNet-101	35.5	56.5	38.2	20.4	38.7	45.6
SSD300 [14]	VGG-16	25.1	43.1	25.8	6.6	25.9	41.4
SSD512 [14]	VGG-16	28.8	48.5	30.3	10.9	31.8	43.5
DSSD321 [37]	ResNet-101	28.0	46.1	29.2	7.4	28.1	47.6
DSSD513 [37]	ResNet-101	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [18]	ResNet-101-FPN	37.8	57.5	40.8	20.2	41.1	49.2
RefineDet512+ [19]	ResNet-101	41.8	62.9	45.7	25.6	45.1	54.1
CornerNet511 [20]	Hourglass-104	40.5	56.5	43.1	19.4	42.7	53.9
CenterNet511 [38]	Hourglass-104	44.9	62.4	48.1	25.6	47.4	57.4
EfficientDet-B3 [39]	BiFPN	**47.5**	**66.2**	**51.5**	27.9	**51.4**	**62.0**
Ours	ResNet-50	34.8	55.2	37.6	23.5	44.2	35.7
Ours	ResNet-50-FPN	37.3	58.0	40.6	28.4	44.2	36.4
Ours	ResNet-101	37.9	58.0	40.7	26.3	47.5	40.4
Ours	ResNet-101-FPN	39.2	59.7	43.0	**28.8**	46.8	38.8

Specific Small Object Detection Models

Methods	Backbone	Avg. Precision, IoU:			Avg. Precision, Area:		
		0.5:0.95	0.5	0.75	S	M	L
ION [41]	VGG-16	30.7	52.9	31.7	11.8	32.8	44.8
MultiPath [11]	VGG-16	33.2	51.9	36.3	13.6	37.2	47.8
SOD-MTGAN [7]	ResNet-101	**41.4**	**63.2**	**45.4**	24.7	44.2	**52.6**
Ours	ResNet-101-FPN	39.2	59.7	43.0	**28.8**	**46.8**	38.8

Ablation Study



Method	text	Avg. Precision, IoU:			Avg. Precision, Area:		
		0.5:0.95	0.5	0.75	S	M	L
w/o reconstruction module	ResNet-50	33.2	53.8	35.7	12.0	30.2	34.9
w/o reconstruction module	Res50-FPN	35.7	56.8	38.4	15.1	30.0	35.3
w/o advloss	ResNet-50	32.5	52.6	34.6	19.8	40.1	28.6
w/o advloss	Res50-FPN	35.2	56.2	37.9	23.9	39.1	28.5
w/o clsloss	ResNet-50	32.5	52.8	34.5	1.9	11.5	29.9
w/o clsloss	Res50-FPN	35.1	55.9	38.0	5.2	14.6	31.0
Ours	ResNet-50	32.5	52.8	34.5	19.8	39.8	28.9
Ours	Res50-FPN	35.1	55.9	38.0	23.7	38.7	28.6

Qualitative Results

Contents

Conclusion

- Generative and discriminative learning framework
- Reconstruction module for anchor box identification
- State-of-the-art accuracy on small object detection

Thanks!