ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

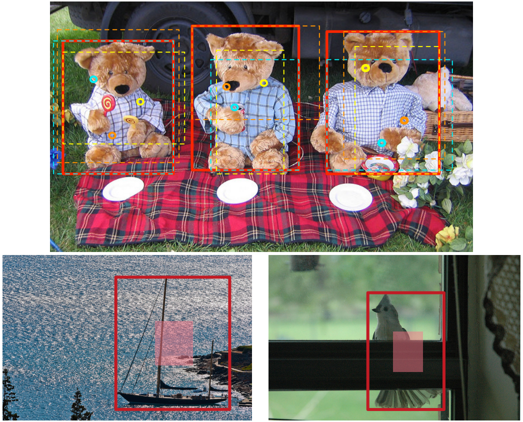
ECCV'20 ONLINE
23-28 AUGUST 2020

Dive Deeper Into Box for Object Detection

Ran Chen[1], Yong Liu[2], Mengdan Zhang[2], Shu Liu[3], Bei Yu[1], Yu-Wing Tai[4]

[1]The Chinese University of Hong Kong
[2]Tencent Youtu Lab
[3]SmartMore
[4]The Hong Kong University of Science and Technology

Introduction



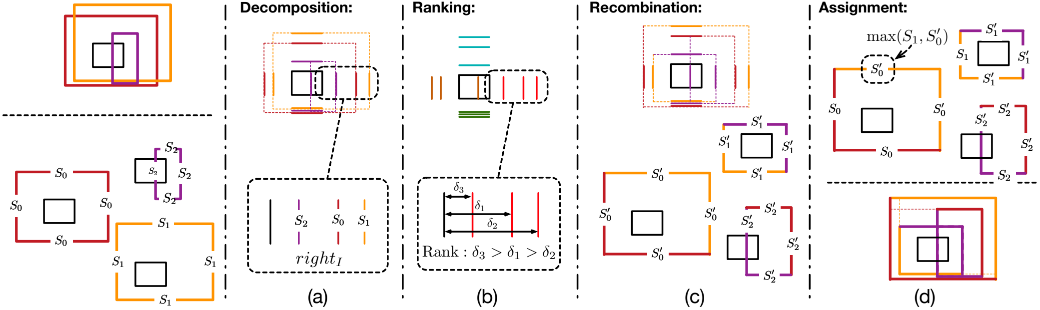
Box Decomposition and Recombination

- Reorganizing boundaries of boxes during training.
- Optimal boxes with tightening instances provide better localization.

Semantic inconsistency in annotations

- Backgrounds regarded as positive pixels are the noise for the training.
- A self adaptive module is approached to tackle this problem.

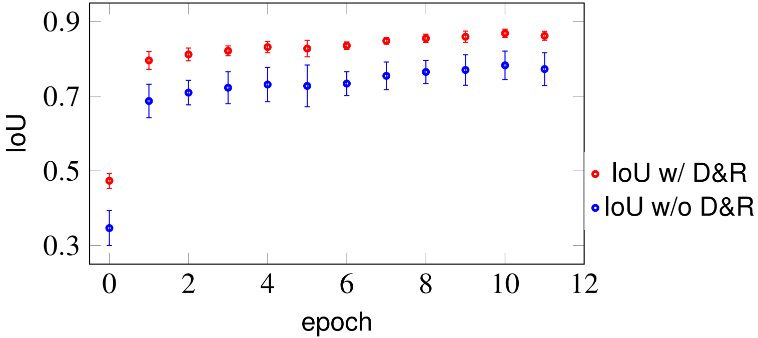
Box Decomposition and Recombination



$$L_{IoU} = -\frac{1}{N_{pos}} \sum_{I} \sum_{i}^{n} \log(IoU(p_i, p_I^*)), \quad (1)$$

$$L_{IoU}^{D\&R} = \frac{1}{N_{pos}} \sum_{I} (\mathbb{1}_{\{S_I' > S_I\}} L_{IoU}(B_I', T_I) + \mathbb{1}_{\{S_I \geqslant S_I'\}} L_{IoU}(B_I, T_I)), \quad (2)$$

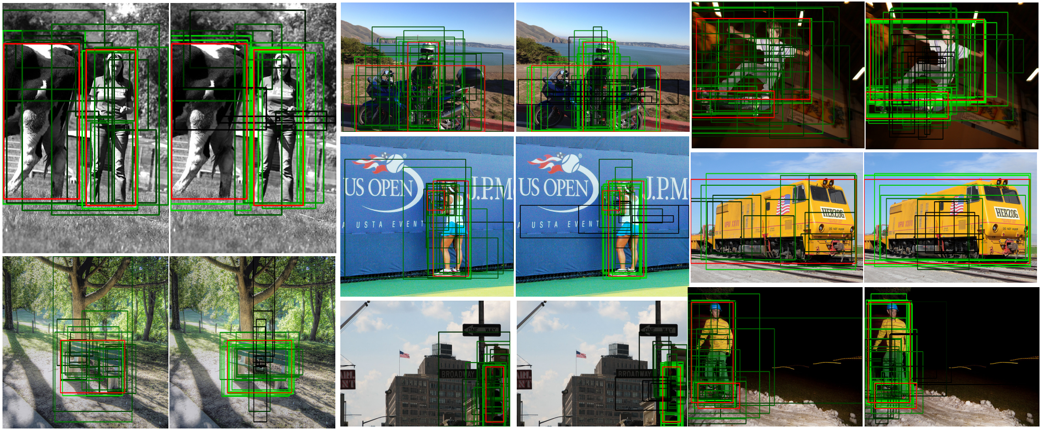
Box Decomposition and Recombination



- Boxes optimized by D&R have higher IoU scores and lower variances.

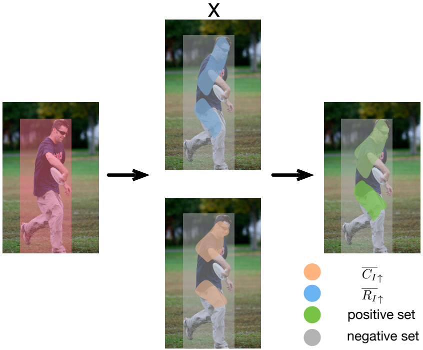
ECCV'20 ONLINE 23-28 AUGUST 2020

Box Decomposition and Recombination



ECCV'20 ONLINE 23-28 AUGUST 2020

Semantic Consistency



$$\begin{cases} \overline{C_{I\downarrow}} \bigcap \overline{R_{I\downarrow}} \leftarrow \text{negative}, \\ \overline{C_{I\uparrow}} \bigcup \overline{R_{I\uparrow}} \leftarrow \text{positive}, \end{cases} \tag{3}$$

$$c_i = \max_{j=0}^{g}(c_j) \in C_I,$$

Settings	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
None	33.6	53.1	35.0	18.9	38.2	43.7
PN	34.2	53.2	36.3	20.8	38.9	44.2
PNI	33.7	53.0	35.5	17.9	38.3	44.1
Ours	**35.3**	**55.4**	**37.1**	**20.9**	**39.6**	**45.9**

Semantic Consistency

Results

Modules			AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Baseline	D&R	Consistency						
✓			33.6	53.1	35.0	18.9	38.2	43.7
✓	✓		34.8	54.0	36.4	19.7	39.0	44.9
✓		✓	37.2	55.4	39.5	21.0	41.7	48.6
✓	✓	✓	**38.0**	**56.5**	**40.8**	**21.6**	**42.4**	**50.4**

Results

Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Two-stage methods:							
Faster R-CNN w/ FPN	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN w/ TDM	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	52.1
Faster R-CNN by G-RMI	Inception-ResNet-v2	34.7	55.5	36.7	13.5	38.1	52.0
RPDet	ResNet-101-DCN	42.8	65.0	46.3	24.9	46.2	54.7
Cascade R-CNN	ResNet-101	42.8	62.1	46.3	23.7	45.5	55.2
One-stage methods:							
YOLOv2	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5
SSD	ResNet-101	31.2	50.4	33.3	10.2	34.5	49.8
DSSD	ResNet-101	33.2	53.3	35.2	13.0	35.4	51.1
FSAF	ResNet-101	40.9	61.5	44.0	24.0	44.2	51.3
RetinaNet	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	53.9
CornerNet	Hourglass-104	40.5	56.5	43.1	19.4	42.7	53.9
ExtremeNet	Hourglass-104	40.1	55.3	43.2	20.3	43.2	53.1
FCOS†	ResNet-101-FPN	41.5	60.7	45.0	24.4	44.8	51.6
FCOS†	ResNeXt-64x4d-101-FPN	43.2	62.8	46.6	26.5	46.2	53.3
FCOS† w/improvements	ResNeXt-64x4d-101-FPN	44.7	64.1	48.4	27.6	47.5	55.6
DDBNet (Ours)	ResNet-101-FPN	42.0	61.0	45.1	24.2	45.0	53.3
DDBNet (Ours)	ResNeXt-64x4d-101-FPN	43.9	63.1	46.7	26.3	46.5	55.1
DDBNet (Ours)§	ResNeXt-64x4d-101-FPN	**45.5**	**64.5**	**48.5**	**27.8**	**47.7**	**57.1**