DualBox: Generating BBox Pair with Strong Correspondence via Occlusion Pattern Clustering and Proposal Refinement

Zheng Ge, Chuyu Hu, Xin Huang, Baiqiao Qiu, Osamu Yoshie --Waseda University

Introduction

➢ Problem:

- The heavy occlusion between pedestrians imposes great challenges to the standard NMS. A relative low threshold of IoU leads to missing highly overlapped pedestrians, while a higher one brings in plenty of false positives.
- R2NMS was proposed to solve the above problem, but it requires complex assigning strategy between anchor pairs and ground-truth boxes.

➢ Novelty:

1. A brand-new pedestrian detector is proposed to simultaneously generate full and visible BBoxes pairs of single pedestrian.
2. We incorporate clustering analysis on occlusion patterns into the learning of visible body proposals and propose a novel occlusion branch to reduce potential visible body proposal.
3. Our DualBox is well-compatible with some famous variants of NMS (e.g., R2NMS, Joint NMS)

➢ Achievement:

- SOTA results but a simpler method compared to PBM in R2NMS on CrowdHuman and Citypersons datasets.

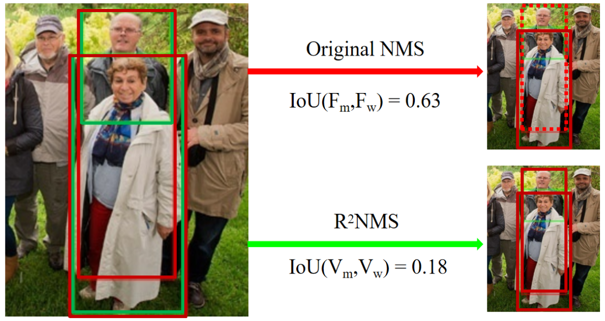
Contect

Email: jokerzz@fuji.waseda.jp
chuyuhu@fuji.waseda.jp

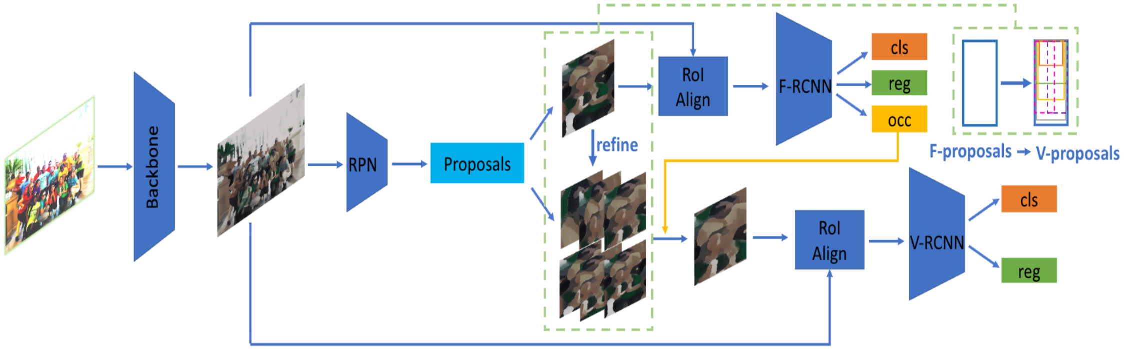
Method

➢ Revisiting R2NMS:

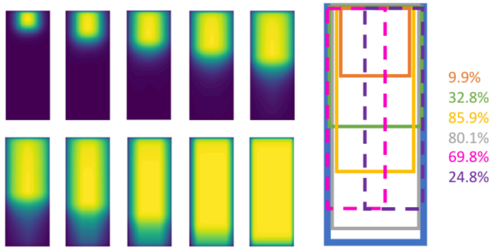
Red BBoxes are full body predictions and green BBoxes are visible body predictions. Two small images on the right show the final results which is processed by original NMS and R2NMS. The red solid BBox represents the preserved BBoxes while red dotted BBox indicates the reduced true positive BBox. The IoU of their full body prediction is 0.63 while the IoU of their visible body is only 0.18. Thus, original NMS will reduce the red dotted BBox but R2NMS is able to keep it.



➢ DualBox:



➢ Clustered Occlusion Patterns:



➢ Occlusion Pattern Clustering:

- **V-Ratio:** We define V-Ratio as $\left(\frac{xv - xf}{wf}, \frac{yv - yf}{hf}, \frac{wv}{wf}, \frac{hv}{hf}\right)$. The clustering algorithm is performed on V-Ratios calculated on ground-truth annotations.
- **V-Box Estimation:** One can estimate visible body proposals from predicted full body proposals $x_v' = x_f' + \Delta x * w_f, w_v' = w_f * \Delta w$ V-Ratio (Occlusior $y_v = y_f + \Delta y * h_f, h_v = h_f * \Delta h$

Experimental Results

➢ Main Results:

Performances on the CrowdHuman val subset. * stands for our re-implemented Faster R-CNN baseline. FV-RCNN and Refine FV-RCNN are two intermediate model before our proposed DualBox. MR_v MR, AP and Recall are reported for evaluation. As shown in this table, our DualBox achieves the best performance.

Method	MR_V	MR	AP	Recall	ΔMR_V	ΔMR
Baseline [2]	55.94	50.42	84.95	90.24		
Baseline*	54.67	47.64	83.79	87.86		
FV-RCNN	55.41	46.32	84.62	88.35		+1.32
Refine FV-RCNN	53.61	46.55	84.74	88.36	+1.06	+1.09
DualBox	**53.25**	**45.65**	84.82	88.38	**+1.42**	**+1.99**

➢ Compatibility with some Variants of NMS:

Method	MR	AP	Recall
NMS	45.65	84.82	88.38
R^2NMS [1]	45.34	86.27	91.33
Joint NMS [6]	45.89	85.81	92.20

➢ State-of-the-art Comparison on CityPersons:

- We list the performance of previous works on reasonable subsets with the original input size in the Table

Methods	R^2NMS	Joint NMS	**R**	**HO**
Baseline (MGAN) [16]			13.8	57.0
Baseline*			13.7	58.3
DualBox			11.5	54.7
DualBox	✓		11.4	54.2
DualBox		✓	11.4	54.3

References

- X. Huang, Z. Ge, Z. Jie, and O. Yoshie, “NMS by representative region: Towards crowded pedestrian detection by proposal pairing,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10 750–10 759.Shuai
- Shao, Zijian Zhao, Boxun Li, Tete Xiao, Gang Yu,Xiangyu Zhang, and Jian Sun. Crowdhuman: A bench-mark for detecting human in a crowd.arXiv preprintarXiv:1805.00123, 2018.