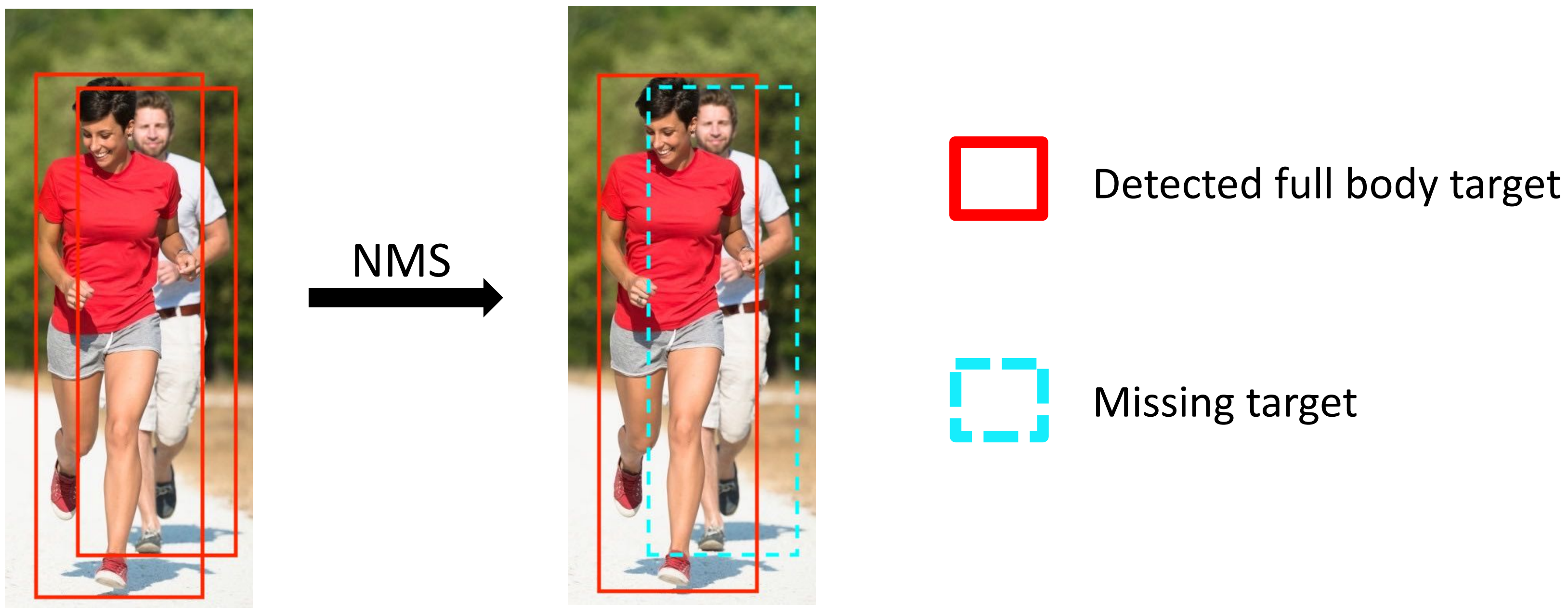
DualBox: Generating BBox Pair with Strong Correspondence via Occlusion Pattern Clustering and Proposal Refinement

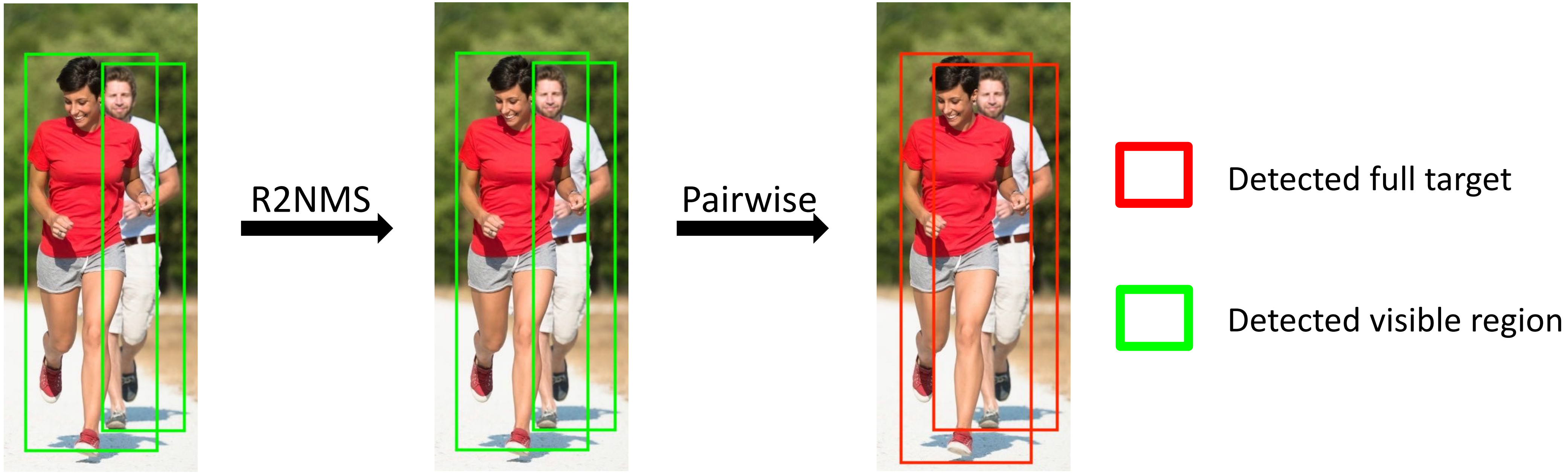
Zheng Ge, Chuyu Hu, Xin Huang, Baiqiao Qiu, Osamu Yoshie
Graduate School of Information, Production and Systems, Waseda University, Japan

Introduction



NMS result: under heavily occluded condition, some true positives are wrongly suppressed.

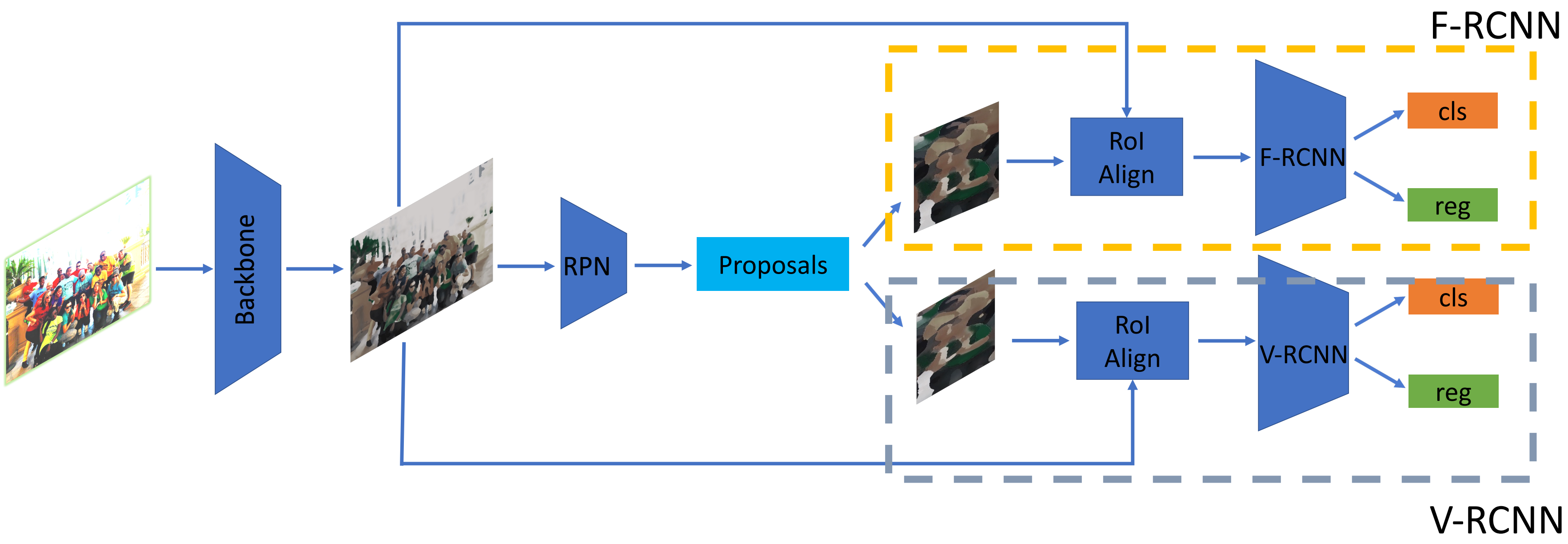
R2NMS[1]



R2NMS result: use visible region BBoxes to calculate IoU and do NMS. Then replace them with the pairwise full body BBoxes.

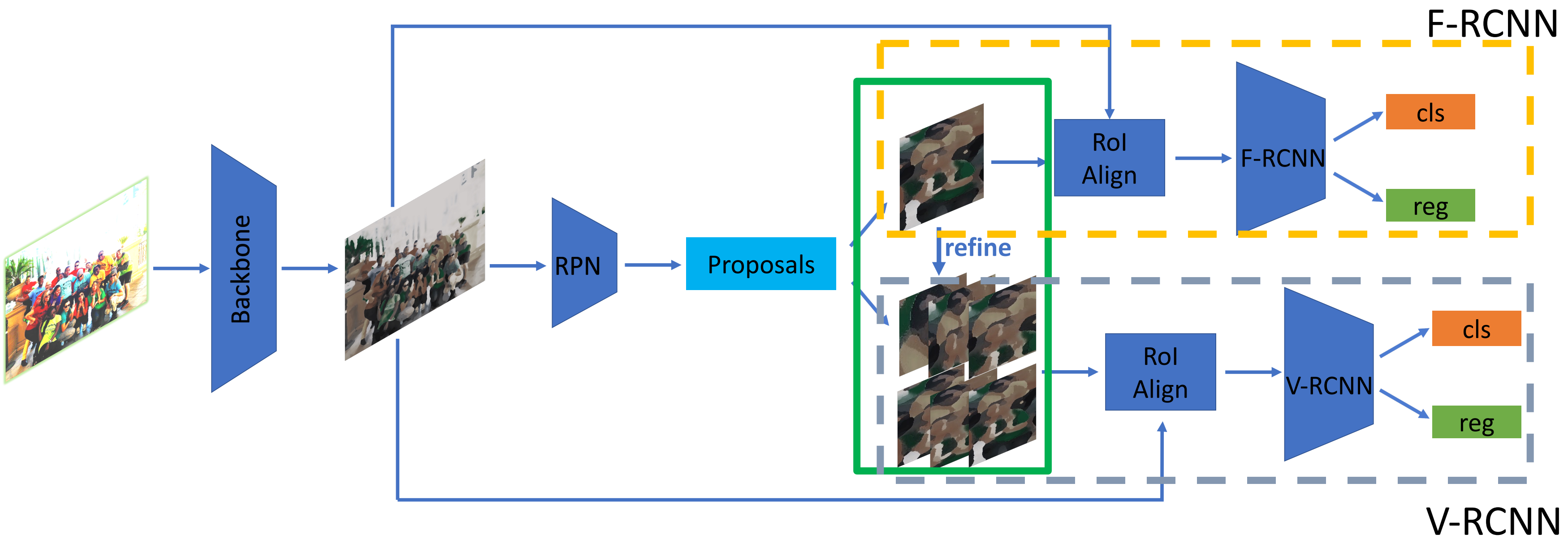
[1] X. Huang, Z. Ge, Z. Jie, and O. Yoshie, “Nms by representative region: Towards crowded pedestrian detection by proposal pairing,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 750–10 759.

DualBox ---- Fundemental architecture



FV-RCNN architecture

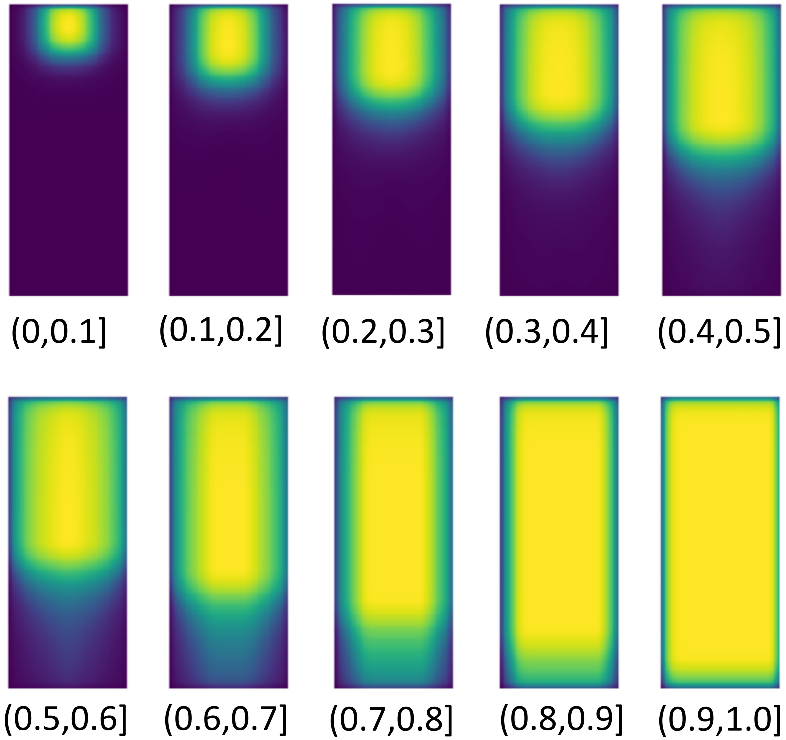
DualBox ---- Assigning strategy



Refine FV-RCNN architecture

Several occlusion patterns are adpoted on each full body proposal to refine the full body proposals into visible body proposals without complex assigning strategies.

DualBox ---- Occlusion patterns



Distribution of occlusion patterns under different degrees of visibility

Full body box $\xrightarrow{\text{V-ratio}}$ **Visible region box**

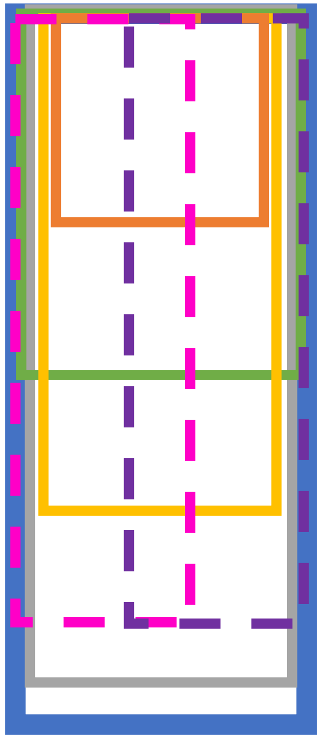
DualBox ---- Occlusion patterns

Different candidate occlusion patterns are implemented on all the full body annotations.

Calculate the IoU between new boxes and original visible body annotations.

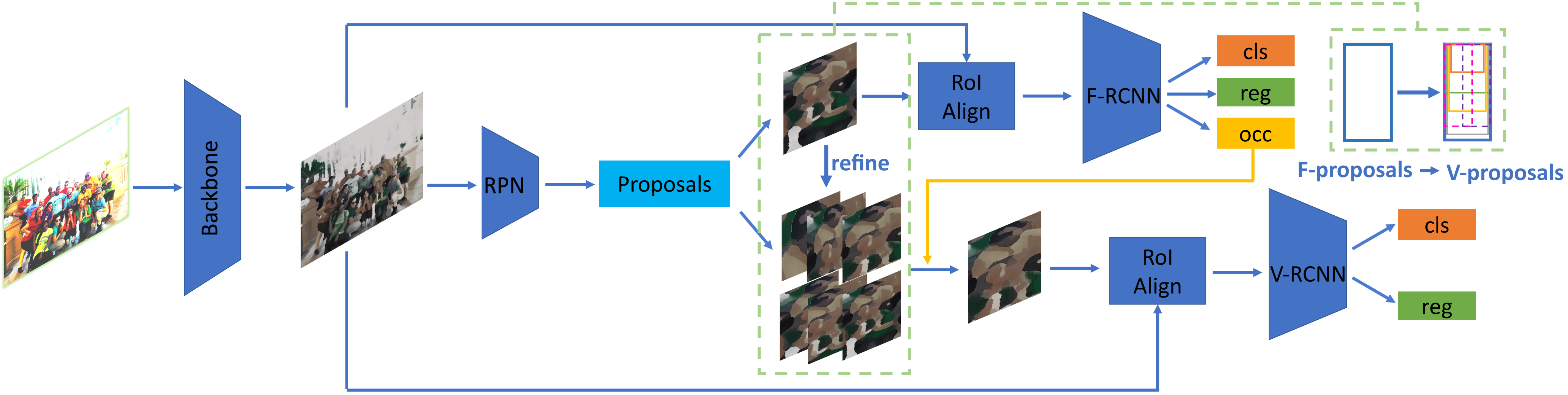
If IoU > 0.5: match
Otherwise: not match

matching rate = # matching boxes / # total boxes



9.9% Head
32.8% Half body
85.9% Most half body
80.1% Almost full body
69.8% Left body
24.8% Right body

DualBox ---- Occlusion branch and final architecture



DualBox architecture

Experiments on CrowdHuman[2]

Method	MR_V	MR	AP	Recall	ΔMR_V	ΔMR
Baseline	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**

Experimental results on CrowdHuman

[2] S. Shao, Z. Zhao, B. Li, T. Xiao, G. Yu, X. Zhang, and J. Sun, "Crowdhuman: A benchmark for detecting human in a crowd," *arXiv preprint arXiv:1805.00123*, 2018.

Experiments on CityPersons[3]

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

Experimental results on CityPersons

[3] S. Zhang, R. Benenson, and B. Schiele, "Citypersons: A diverse dataset for pedestrian detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3213–3221.

Ablation study

Method	parallel	cascade	MR_V	MR	AP	Recall
Baseline*			54.67	47.64	83.79	87.86
FV-RCNN	√		55.41	46.32	84.62	88.35
		√	58.84	46.39	84.62	88.74
refine FV-RCNN	√		53.61	46.55	84.74	88.36
		√	53.47	46.59	84.59	88.55
DualBox	√		53.25	**45.65**	84.82	88.38
		√	**53.21**	46.11	84.87	88.39

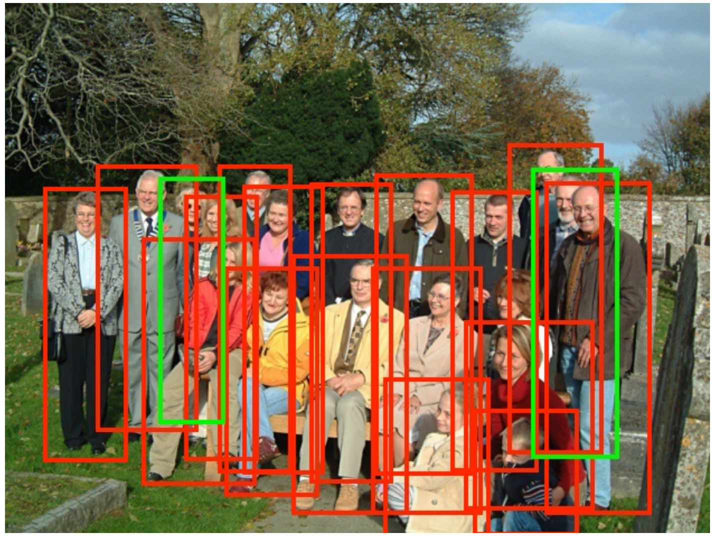
Ablation study about parallel and cascade refinement mode on CrowdHuman

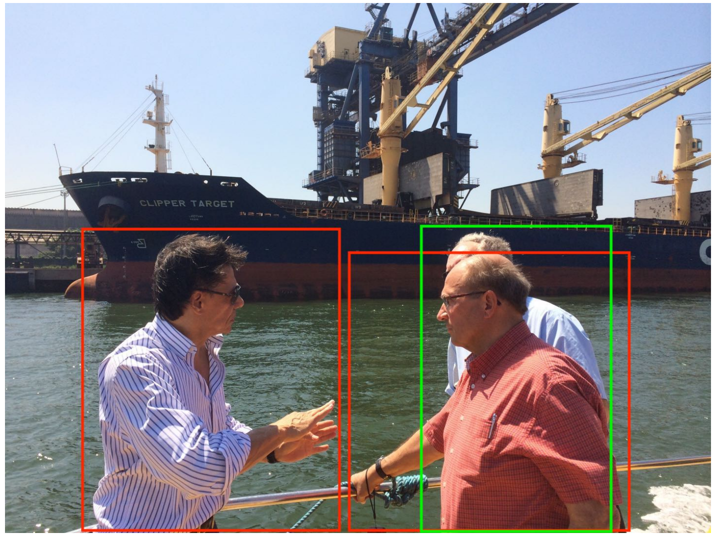
Method	MR	AP	Recall
NMS	45.65	84.82	88.38
R^2NMS	45.34	86.27	91.33

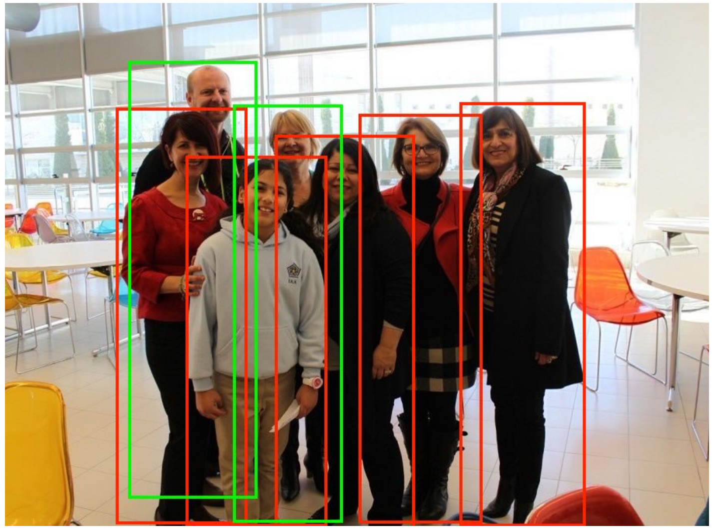
Ablation study about post processing approaches on CrowdHuman

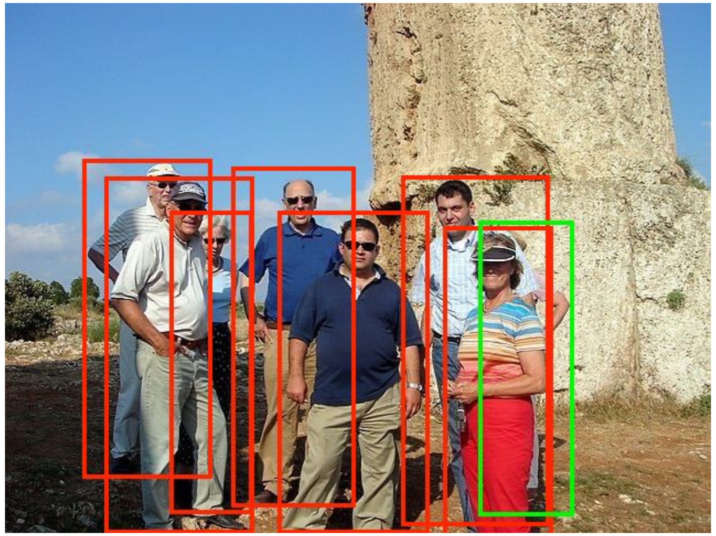
[3] S. Zhang, R. Benenson, and B. Schiele, "Citypersons: A diverse dataset for pedestrian detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3213–3221.

Visulization results









Detected results of baseline Faster R-CNN

Extra detected results of our DualBox

Thanks for listening