COUNTING DENSE OBJECTS IN REMOTE SENSING IMAGES

Guangshuai Gao[1], Qingjie Liu[1,2,], Yunhong Wang[1,2]*

[1] The State Key Laboratory of Virtual Reality Technology and Systems, Beihang University
[2] Hangzhou Innovation Institute, Beihang University, Hangzhou China

IRIP laboratory

Main challenge

- Dataset scarcity;
- Scale variation;
- Complex cluttered background;
- Orientation arbitrariness

Contribution

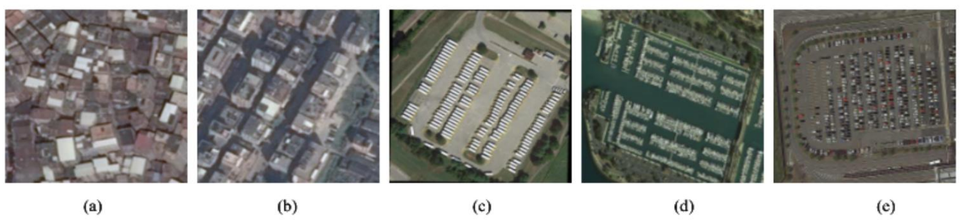

Fig. 1: Representative images of RSOC dataset. (a)-(e) represent the images of building_A, building_B, large-vehicle, ship and small-vehicle, respectively.

Table 1: There are four different kinds of objects in the dataset: building, ship, large vehicle and small vehicle. We split the dataset into five subsets according to the density levels of the objects. Building_A has larger density level than Building_B. It also worth noting that for data annotation, building subsets are labeled with center point, and other three ones adopt bounding box, preprocessing to compute their center points when generating their ground truth density maps.

Dataset	Images	Training/Test	Average Resolution	Statistics			
				Total	Min	Average	Max
Building_A	360	179/181	512 × 512	19,963	34	55.45	142
Building_B	2108	1026/1082	512 × 512	56,252	15	26.69	76
Ship	137	97/40	2558 × 2668	44,892	50	327.68	1661
Large-vehicle	172	108/64	1552 × 1573	16,594	12	96.48	1336
Small-vehicle	280	222/58	2473 × 2339	148,838	17	531.56	8531

*Correspondence to qingjie.liu@buaa.edu.cn

Method

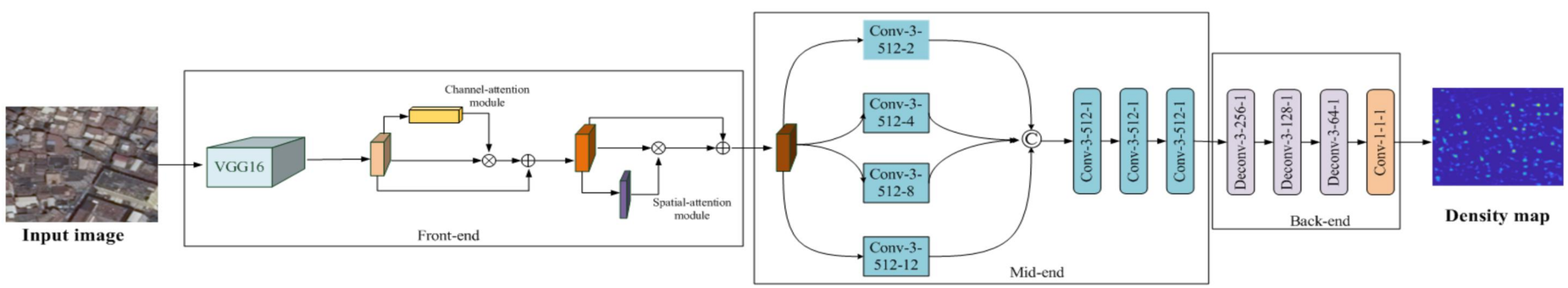

Fig. 2: Architecture of ASPDNet. The parameters of convolutional layers in the mid-end stage are denoted as "Conv-(kernel size)-(number of filters)-(dilation rate)", while the parameters in the back-end stage are represented as "Deconv-(kernel size)-(number of filters)-(stride)". The max-pooling layers are conducted over a 2 × 2 pixel window, with stride 2 (ignored in the figure). "⊕", "⊗" and "ⓒ" represent matrix addition, matrix multiplication and matrix concatenation operation, respectively.

$$L(\Theta) = \frac{1}{2N}\sum_{i=1}^{N}\left\|F(X_i;\Theta) - F_i^{GT}\right\|_2^2$$

Results

Table 2: Performance comparison on Building_A and Building_B dataset.

Method	Building_A		Building_B	
	MAE	RMSE	MAE	RMSE
MCNN [1]	14.33	19.47	13.11	16.60
CMTL [3]	15.04	20.77	10.24	13.64
CSRNet [2]	13.32	18.00	7.37	10.32
SFCN [4]	13.14	17.56	6.31	8.33
ASPDNet(our proposed)	**10.21**	**14.27**	**5.31**	**7.02**

Table 4: Ablation studies on Building_A and Building_B dataset.

Method	Building_A		Building_B	
	MAE	RMSE	MAE	RMSE
Baseline	13.32	18.00	7.37	10.32
Baseline+CBAM	12.95	17.51	7.71	9.90
Baseline+CBAM+SPM	12.28	16.52	5.67	7.28
ASPDNet	**10.21**	**14.27**	**5.31**	**7.02**

Table 3: Performance comparison on Ship-, Large-vehicle and Small-vehicle sub datasets.

Method	Ship		Large-vehicle		Small-vehicle	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
MCNN [1]	263.91	412.30	36.56	55.55	488.65	1317.44
CMTL [3]	251.17	403.07	61.02	78.25	490.53	1321.11
CSRNet [2]	240.01	394.16	34.10	46.42	443.72	1252.22
SFCN [4]	240.16	394.81	33.93	49.74	440.70	1248.27
ASPDNet(our proposed)	**193.83**	**318.95**	**31.76**	**40.14**	**433.23**	**1238.61**

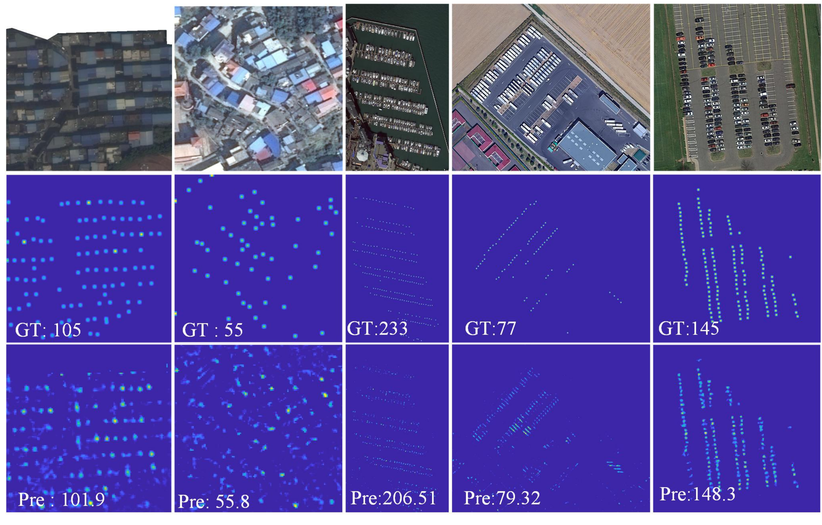

Fig. 3: Example Results on the RSOC dataset with the proposed method.