Few-Shot Scene Adaptive Crowd Counting Using Meta-Learning
Supplementary Material

Mahesh Kumar Krishna Reddy[1] Mohammed Asiful Hossain[2] Mrigank Rochan[1] Yang Wang[1]

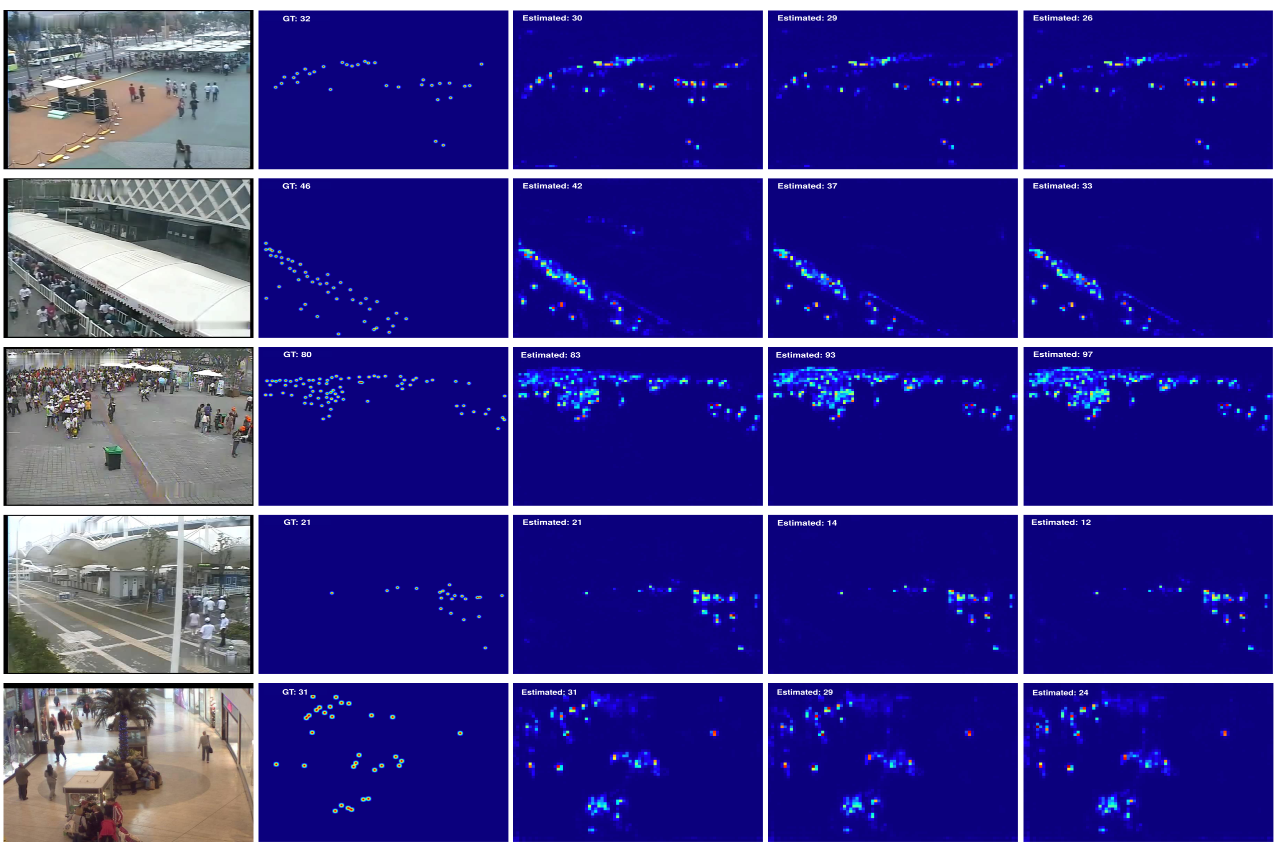
[1]University of Manitoba [2]Huawei Technologies Co., Ltd.

{kumarkm, mrochan, ywang}@cs.umanitoba.ca asif07hossain@gmail.com

In this supplementary material, we show the results of the qualitative evaluation in Sec. 1 and in Sec. 2 we present the quantitative results on WorlExpo'10 [4] test set for different meta-learning approaches.

1. Qualitative Evaluation

We show some density maps generated by our method and baselines in Fig. 1. In general, our method produces density maps with estimated crowd count closer to the ground-truth than the baselines.



(a) Input Image (b) Ground-truth (c) Ours (d) Baseline fine-tuned (e) Baseline pre-trained

Figure 1. Qualitative results showing the generated density maps for different models. Here we include, (a) input scene-specific crowd image, (b) the corresponding ground-truth (GT), (c) predictions (estimated count) from our proposed approach, (d) predictions from fine-tuned baseline and (e) predictions from pre-trained baseline. The first four rows are the test scenes from WorldExpo'10 [4] and the last row is the test scene from Mall [1] dataset.

2. Quantitative Evaluation

We show some quantitative evaluation performance of different optimization based meta-learning approaches [3, 2] on different scenes in WorldExpo test set.

Target	Methods	1-shot (K=1)			5-shot (K=5)		
		MAE	RMSE	MDE	MAE	RMSE	MDE
Scene 1	Meta-LSTM [3]	4.52 ± 1.06	5.95 ± 1.62	0.40 ± 0.08	3.66 ± 0.85	4.64 ± 1.30	0.39 ± 0.10
	Reptile [2]	4.99 ± 0.46	7.08 ± 1.09	0.47 ± 0.123	3.38 ± 0.59	4.38 ± 0.63	0.36 ± 0.112
	Ours w/o ROI	3.47 ± 0.01	**4.19 ± 0.01**	0.50 ± 0.007	3.42 ± 0.03	4.81 ± 0.007	**0.29 ± 0.004**
	Ours w/ ROI	**3.19 ± 0.03**	4.30 ± 0.07	**0.38 ± 0.03**	**3.05 ± 0.06**	**4.19 ± 0.15**	0.31 ± 0.08
Scene 2	Meta-LSTM [3]	19.09 ± 3.54	26.42 ± 5.11	0.22 ± 0.01	18.89 ± 1.87	26.35 ± 2.61	0.160 ± 0.014
	Reptile [2]	21.51 ± 0.45	25.85 ± 0.48	0.30 ± 0.062	14.52 ± 3.11	20.46 ± 4.29	0.14 ± 0.049
	Ours w/o ROI	12.05 ± 0.74	16.62 ± 1.10	0.11 ± 0.007	11.41 ± 0.54	15.35 ± 0.51	0.11 ± 0.015
	Ours w/ ROI	**11.17 ± 1.01**	**15.50 ± 1.18**	**0.11 ± 0.012**	**10.73 ± 0.36**	**14.95 ± 0.60**	**0.10 ± 0.003**
Scene 3	Meta-LSTM [3]	24.66 ± 1.13	33.54 ± 1.22	0.29 ± 0.011	24.24 ± 0.95	29.38 ± 1.15	0.27 ± 0.008
	Reptile [2]	14.14 ± 1.68	18.04 ± 1.71	0.156 ± 0.015	9.37 ± 1.30	12.04 ± 1.27	0.123 ± 0.026
	Ours w/o ROI	8.15 ± 0.17	11.04 ± 0.42	**0.09 ± 0.04**	8.31 ± 0.54	**10.75 ± 0.54**	0.10 ± 0.009
	Ours w/ ROI	**8.07 ± 0.23**	**10.92 ± 0.21**	0.10 ± 0.007	**8.18 ± 0.24**	10.96 ± 0.31	**0.09 ± 0.002**
Scene 4	Meta-LSTM [3]	12.88 ± 0.88	15.83 ± 0.9	0.114 ± 0.007	12.03 ± 0.22	14.78 ± 0.36	0.105 ± 0.002
	Reptile [2]	12.09 ± 1.28	14.64 ± 1.09	0.106 ± 0.017	9.64 ± 1.22	12.11 ± 1.34	0.082 ± 0.011
	Ours w/o ROI	9.74 ± 0.09	11.9 ± 0.12	0.084 ± 0.001	11.21 ± 0.47	16.1 ± 0.45	0.118 ± 0.004
	Ours w/ ROI	**9.39 ± 0.26**	**11.78 ± 0.34**	**0.07 ± 0.02**	**9.41 ± 0.21**	**11.91 ± 0.17**	**0.08 ± 0.002**
Scene 5	Meta-LSTM [3]	5.54 ± 0.29	9.36 ± 0.35	0.24 ± 0.01	4.68 ± 0.17	7.92 ± 0.577	0.194 ± 0.002
	Reptile [2]	5.42 ± 0.32	9.75 ± 0.34	0.27 ± 0.081	4.10 ± 0.66	7.57 ± 1.20	0.204 ± 0.058
	Ours w/o ROI	4.09 ± 0.01	7.36 ± 0.01	0.196 ± 0.001	4.28 ± 0.14	7.68 ± 0.60	0.20 ± 0.001
	Ours w/ ROI	**3.82 ± 0.05**	**6.91 ± 0.11**	**0.192 ± 0.001**	**3.91 ± 0.26**	**7.18 ± 0.85**	**0.18 ± 0.001**
Average	Meta-LSTM [3]	13.33	18.22	0.252	12.7	16.61	0.223
	Reptile [2]	11.63	15.07	0.260	8.20	11.31	0.181
	Ours w/o ROI	7.5	10.22	0.197	7.7	10.93	0.165
	Ours w/ ROI	**7.12**	**9.88**	**0.172**	**7.05**	**9.83**	**0.155**

Table 1. The overall results for adaptation on WorldExpo'10 [4] test set with $K = 1$ and $K = 5$ train images. We explore alterntaive optimization based meta-learning approaches such as *Meta-LSTM* [3] and *Reptile* [2] along with our models *"Ours w/o ROI"* and *"Ours w/ ROI"*. The results in bold represent the overall best result.

References

[1] C. C. Loy, S. Gong, and T. Xiang. From semi-supervised to transfer counting of crowds. In *IEEE International Conference on Computer Vision*, 2013.

[2] A. Nichol, J. Achiam, and J. Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.

[3] S. Ravi and H. Larochelle. Optimization as a model for few-shot learning. In *International Conference on Learning Representations*, 2017.

[4] C. Zhang, H. Li, X. Wang, and X. Yang. Cross-scene crowd counting via deep convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.