Weighing Counts: Sequential Crowd Counting by Reinforcement Learning Supplement Materials

Liang Liu[1], Hao Lu[2], Hongwei Zou[1], Haipeng Xiong[1], Zhiguo Cao[1], and Chunhua Shen[2]

[1] School of Aritifical Intelligence & Automation, Huazhong University of Science and Technology, China
[2] The University of Adelaide, Australia
{wings,zgcao}@hust.edu.cn

1 Discretization and Inverse-Quantization

In this section, we illustrate the generation of counting intervals ('Discretization') and inverse-quantization [2] in detail.

1.1 Counting Intervals Generation

First, given a map of dotted annotations, it is convoluted by the Gaussian kernel to compute the density map $\mathcal{D}(p)$ [1], which takes the form

$$\mathcal{D}(p) = \sum_{i=1}^{N} \delta(p - D_i) * G_{\sigma_i}(p) \tag{1}$$

where $p \in I$ is a pixel in the image I, D_i is the i-th dot annotation of I, and G_{σ_i} is a Gaussian kernel with the variance of σ_i. In this paper, we employ the adaptive Gaussian kernel [4], whose variance is defined by

$$\sigma_i = \beta \bar{d}_i \, , \tag{2}$$

where $\bar{d}_i$ is the average distance between the dot point D_i and its 3-nearest dot points, β is a hyperparameter which is set to 0.3 following [4].

Further, the density map is summed at patch-level to compute patch count $\mathcal{N}$ [3]

$$\mathcal{N}_i = \sum_{p_b \in P_i} \mathcal{D}(p_b) \, , \tag{3}$$

where P_i is the i-th patch of the image I.

Finally, the count value is quantized to compute the count interval $\mathcal{C}$ [2]

$$\mathcal{C}_i = \mathcal{Q}(\mathcal{N}_i) = \begin{cases} 0 & \mathcal{N}_i = 0 \\ max\Big(floor\Big(\frac{log(\mathcal{N}_i) - l}{w} + 2\Big), 1\Big) & Otherwise \end{cases} , \tag{4}$$

where w is the width of the quantized interval in log space, l is a hyperparameter which means the interval of $\left(0, e^l\right]$ is divided as an independent class [2]. In this paper, we set $w = 0.1$ and $l = -2$.

1.2 Inverse-Quantization

During testing, the counting value is recovered from the counting interval by the inverse-quantization $\mathcal{IQ}$ [2], i.e.,

$$\mathcal{N}_i = \mathcal{IQ}(\mathcal{C}_i) = \begin{cases} 0 & \mathcal{C}_i = 0 \\ \frac{1}{2}exp(l + w(\mathcal{C}_i - 1)) & \mathcal{C}_i = 1 \\ \frac{1}{2}exp(l + w(\mathcal{C}_i - 2)) + \frac{1}{2}exp(l + w(\mathcal{C}_i - 1)) & Otherwise \end{cases} \tag{5}$$

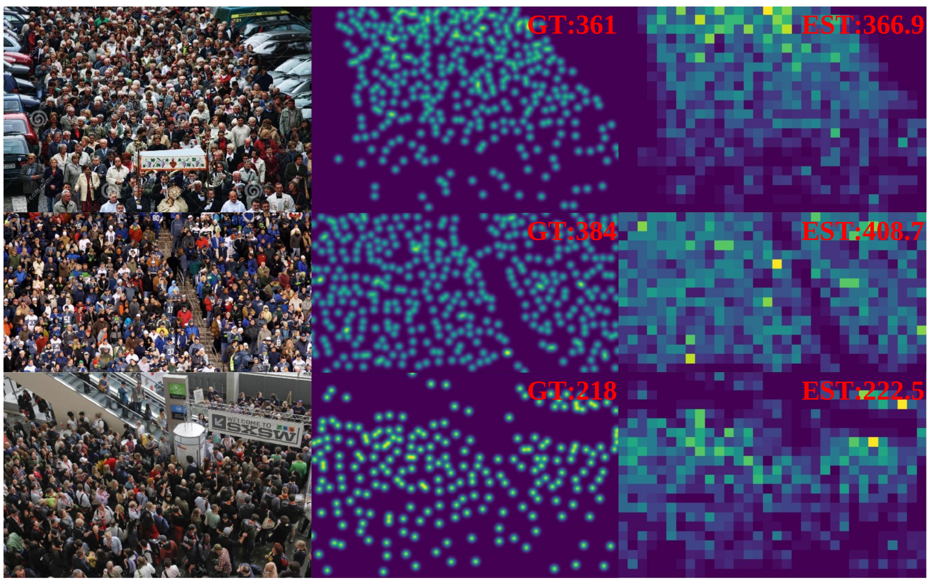

Fig. I. Qualitative results of LibraNet on the ShanghaiTech part_A dataset. From left to right, there are testing images, density maps with ground-truth counts, and our estimated results.

References

1. Lempitsky, V., Zisserman, A.: Learning to count objects in images. In: Advances in Neural Information Processing Systems (NIPS). pp. 1324–1332 (2010)
2. Liu, L., Lu, H., Xiong, H., Xian, K., Cao, Z., Shen, C.: Counting objects by blockwise classification. IEEE Transactions on Circuits and Systems for Video Technology (2019)
3. Lu, H., Cao, Z., Xiao, Y., Zhuang, B., Shen, C.: TasselNet: counting maize tassels in the wild via local counts regression network. Plant methods **13**(1), 79 (2017)
4. Zhang, Y., Zhou, D., Chen, S., Gao, S., Ma, Y.: Single-image crowd counting via multi-column convolutional neural network. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 589–597 (2016)

Fig. II. Qualitative results of LibraNet on the ShanghaiTech part_B dataset. From left to right, there are testing images, density maps with ground-truth counts, and our estimated results.

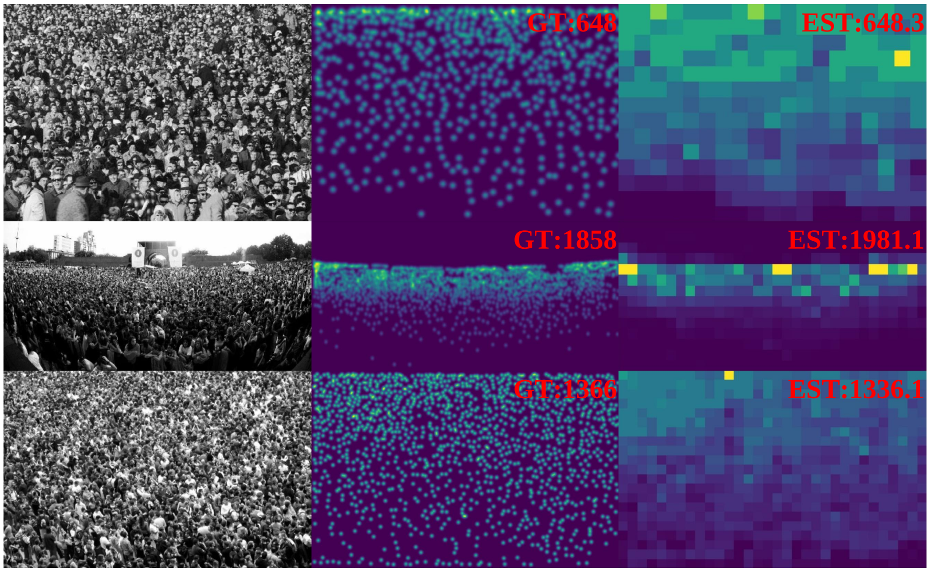

Fig. III. Qualitative results of LibraNet on the UCF_CC_50 dataset. From left to right, there are testing images, density maps with ground-truth counts, and our estimated results.

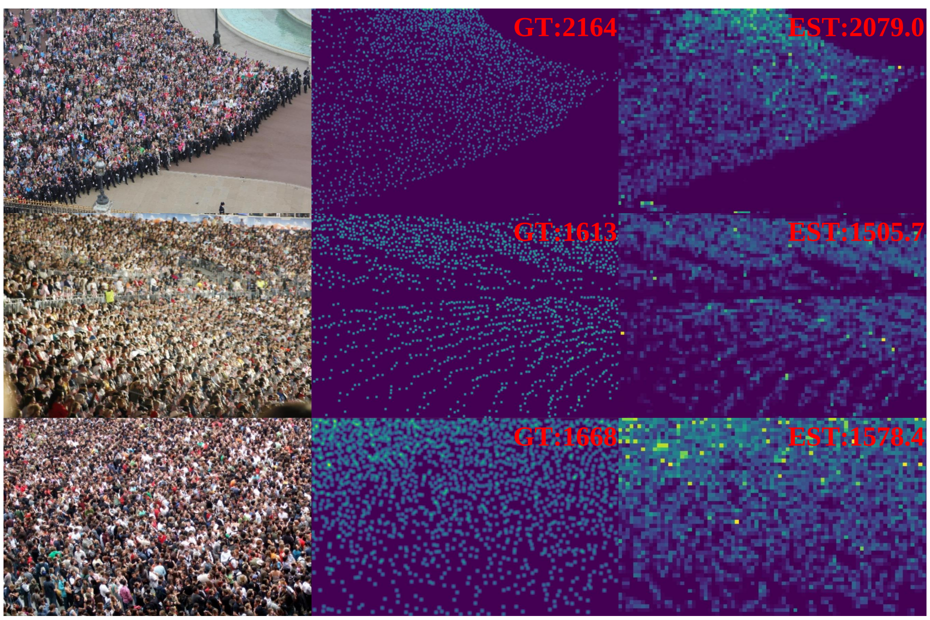

Fig. IV. Qualitative results of LibraNet on the UCF-QNRF dataset. From left to right, there are testing images, density maps with ground-truth counts, and our estimated results.

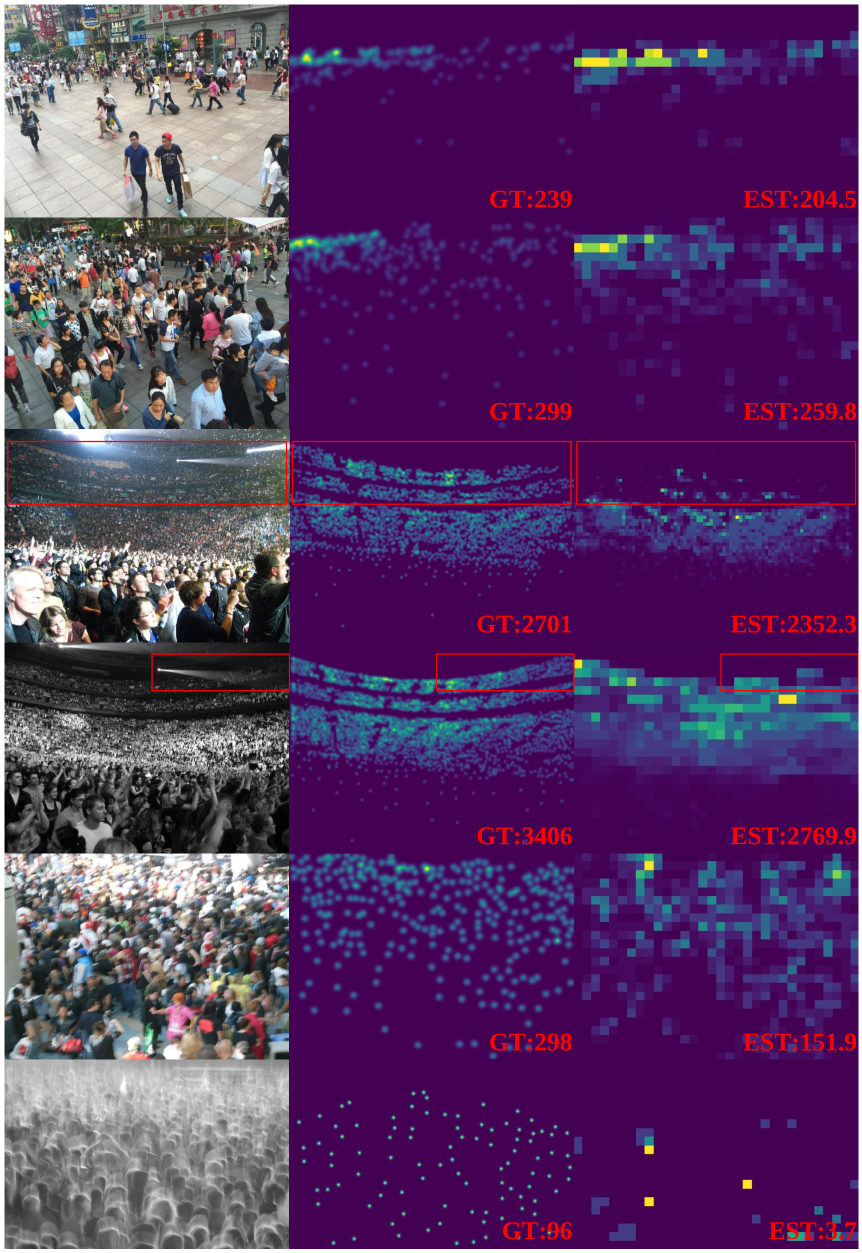

Fig. V. Failure cases. The first 2 cases from the ShanghaiTech Part_B dataset show that our method suffers from the training bias caused by long-tailed distribution, leading to under-estimations. The 3-rd and 4-th rows demonstrate that LibraNet can be affected by illumination variations. The last 2 rows illustrate that our method fails due to blurry appearance. In each row, from left to right, there are testing images, density maps with ground-truth counts, and our estimated results.